

UNIVERSITY OF VERONA



*Department of
Diagnostics and Public Health*

*Doctoral School of
Natural and Engineering Sciences*

*PhD Course in
Nanoscience and Advanced Technologies -
Forensic Sciences and Technologies*

XXXVII Cycle (2021-2024)

***Inferring the Weight of Evidence from two-person
mixtures when explored by Y-chromosome haplotypes***

S.S.D. MED/43

Coordinator: Prof. Adolfo Speghini

Tutor: Prof.ssa Stefania Turrina

PhD student: Dott.ssa Giulia Soldati

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License, Italy. To read a copy of the licence, visit the web page:

<http://creativecommons.org/licenses/by-nc-nd/3.0/>



Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



NonCommercial — You may not use the material for commercial purposes.



NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

Inferring the Weight of Evidence from two-person mixtures
when explored by Y-chromosome haplotypes

Giulia Soldati
PhD thesis

*A Davide,
dal primo giorno al mio fianco*

CLAIMS AND DISCLAIMER

This doctoral thesis stems from my collaboration with Professor Vincenzo Lorenzo Pascali (VLP).

I exchanged ideas with VLP on Y-Chr mixtures during numerous meetings with him in several formal and informal circumstances since June 2023. Furthermore, I had a stint of intensive work with him from June to July 2023 and from October (2023) to January (2024) at the Department of Healthcare Surveillance and Bioethics UCSC Rome.

He generously shared with me original and unpublished ideas – as well as pieces of software/worksheets containing calculation routines – that he had independently conceived and extensively tested on this topic before this thesis was designed and finalized. I acknowledge VLP's intellectual property and primacy over all procedures this thesis illustrates. I thank him for introducing me to this fascinating field of research, and I hope to have the chance to collaborate with him and progress further in the near future.

I hold myself responsible for any unintentional errors, omissions or misinterpretation of his original ideas that I may have incurred while writing this Doctoral thesis.

ABSTRACT

In recent decades, the improvement in the sensitivity of commercial kits used for the amplification of genetic markers in forensic work, together with the increase in the complexity of the cases dealt with, has led to an increase in the number of mixed traces submitted to the forensic expert for analysis as a model of evidence in molecular investigations aimed at criminal identification.

The interpretation of mixed traces analyzed by autosomal markers has been an intensively studied field, and at least two computational methods are now available to the expert: qualitative (semi-continuous) and quantitative Bayesian (continuous). A good assortment of application software inspired by these principles is also available today, with products that can be purchased commercially or acquired as free shareware (for example, LRmix Studio, EuroForMix, CaseSolver, etc.). Some of these software are surveyed and, to some extent, approved by the International Society of Forensic Genetics.

However, what still makes the theoretical and application landscape on the subject of mixed traces unsatisfactory is the almost total absence of principles and analysis methods that can be applied to the field of markers located on the Y chromosome (Y-Chr).

In this context, the main problems that make theoretical elaboration and interpretative progress complex are as follows:

- The Y-Chr is theoretically inherited en bloc, and the presence of haplotypes makes extrapolating population frequency data that can be integrated into individualization calculations challenging;
- A unique haplotype frequency calculation model does not yet exist;
- The approach to the identification of all possible combinations resulting from a mixed profile has to deal with the construction of pairs, triplets, quadruplets, etc. (in the case of mixtures with two, three, and four contributors, respectively) of haplotypes (and not of single genotypes);
- Many haplotypes generated by an albeit correct haplotype combinatorial process may never have been detected in the population surveys carried out by the laboratories; therefore, their existence may legitimately be questioned;

- Solutions for the deconvolution of Y-STRs (Short Tandem Repeats) mixtures are still practically non-existent.

This work aims to illustrate a computational procedure to extract and evaluate the Power of Evidence (WoE) in biological mixtures corresponding to the *two contributors' model*, studied employing a genetic marker protocol on the Y-Chr.

In this study, the Y-STRs mixtures are represented with genotype permutation universal matrices, which include several columns equivalent to the number of putative contributors and many rows/states comparable to the number of different permutations incorporating the evidence. After the matrices' construction, a calculation method was proposed based on the assumption of measurable quantities, gene frequencies assumed to be all available, and no drop-out. The three independent indices (Peak Height ratio, Mix Ratio, and χ^2) were modelled, and a specific description of how they were calculated is given in the manuscript.

The theoretical development of this work was described following the *two contributors – ten loci* model; practical calculation examples helpful in testing the effectiveness of the method developed were then described using the *two contributors – seven loci* model, due to the low data processing capacity of the personal computers used in this work.

The spreadsheets developed in Microsoft Excel, relating to matrix construction and calculation of the Likelihood Ratio, are user-friendly and accessible for reference upon explicit request to the Authors.

ABBREVIATIONS

aSTRs	Autosomal Short Tandem Repeats
CI	Coincidence Probability
CPI	Combined Probability of Inclusion
DNA	Deoxyribonucleic Acid
E	Evidence
E_{mix}	Mixed Evidence
EPG	Electropherogram
E_{ss}	Single Source Evidence
H_d	Defence Hypothesis
H_p	Prosecution Hypothesis
IBD	Identity by Descent
IBS	Identity by State
ISFG	International Society of Forensic Genetics
LR	Likelihood Ratio
LT	Low Template
MR	Mix Ratio
mRMP	Modified Random Match Probability
MSY	Male-Specific region of the Y-Chr
NGS	Next-Generation Sequencing
NRY	Non-recombining Region of the Y-Chr
<i>P</i>	Genotype frequency
PCR	Polymerase Chain Reaction
PhR	Peak Height Ratio
PM	Person Mixture
POI	Person Of Interest
RFU	Relative Fluorescence Units
RM	Rapidly Mutating
RMNE	Random Man Not Excluded
RMP	Random Match Probability
SNPs	Single Nucleotide Polymorphisms

STRs	Short Tandem Repeats
U	Unknown individual
UME	Unique Mutation Event
WoE	Weight/power of Evidence
Y-Chr	Y Chromosome
YHRD	Y-STR Haplotype Reference Database
χ^2	Chi-Square

TABLE OF CONTENTS

CLAIMS AND DISCLAIMER	4
ABSTRACT	5
ABBREVIATIONS	7
TABLE OF CONTENTS	9
INTRODUCTION	12
1.1 Brief History	13
1.2 Y-Chr polymorphisms	14
1.3 Y-Chr polymorphisms forensic applications	16
REFERENCES	18
DNA MIXTURES	22
2.1 DNA mixtures interpretation (aSTRs)	25
2.1.1 Determining the number of contributors	30
2.1.2 Frequentist approaches	30
2.1.2.1 Random Match Probability (RMP)	30
2.1.2.2 Random Man not Excluded (RMNE)	30
2.1.3 Likelihood Ratio approaches	32
2.1.3.1 Semi-continuous model	33
2.1.3.2 Continuous model	35
2.2 DNA mixtures interpretation (Y-STRs)	37
2.2.1 Y chromosome is inherited in allelic association	37
2.2.2 Lack of a unique model for calculating haplotype frequencies	38
2.2.3 Assembling pairs, triplets, and quadruplets as haplotypes	40
2.2.4 Not all conceivable haplotypes are real	41
2.2.5 No developed strategies	41
2.2.6 Low computing power of personal computers	41
REFERENCES	43
AIM OF THE PROJECT	48
MATERIALS AND METHODS	50
4.1 Male DNA samples	50
4.2 Experimental male DNA mixtures	50
4.3 Male in silico DNA mock mixtures	50

4.4 Construction of the universal matrix	51
4.5 Calculation worksheet	54
4.6 Statistical Analysis	54
4.6.1 Calculation of the MRs	54
4.6.2 Calculation of the $(1-\chi^2)$ values	55
4.6.3 Calculation of intra-individual PhR.....	56
4.6.4 Population probabilities	58
4.6.5 Combining the set of evidence in a likelihood	60
4.6.6 Calculation of the Likelihood Ratio	61
4.7 Building worksheets to enable calculations	61
4.8 Data on which to give proof of principle	62
REFERENCES	64
RESULTS	66
5.1 Male DNA samples	66
5.2 Experimental male DNA mixtures	66
5.3 Male in silico DNA mock mixtures	67
5.4 Construction of the universal matrix	69
5.5 Statistical analysis	69
DISCUSSION	72
REFERENCES	79
GENERAL CONCLUSIONS	81
Appendix A	83

CHAPTER 1

INTRODUCTION	12
1.1 Brief History.....	13
1.2 Y-Chr polymorphisms.....	14
1.3 Y-Chr polymorphisms forensic applications.....	16
REFERENCES	18

INTRODUCTION

The sex chromosomes (Chr) determine an individual's sex: females typically have two X-Chr (X, X), while males have one X and one Y-Chr (X, Y).

Gonosomal STRs are particularly useful in studies related to paternity testing, population genetics, and forensic analysis, similar to autosomal STRs.

For example, in paternity testing, analysis of Short Tandem Repeats on the X or Y-Chr (X-STRs and Y-STRs, respectively) can help confirm the biological relationship between a father and a child by examining the inheritance pattern of the STRs on the X and Y-Chr. These latter are especially valuable when analyzing cases involving male relatives.

The transmission of gonosomal STRs follows the patterns of inheritance associated with the sex chromosomes.

X-Chr Transmission:

From Mother to Offspring → The X-Chr is inherited from both parents. Sons inherit their X-Chr from their mother, while daughters inherit one X-Chr from each parent (one from the mother and the other from the father).

From Father to Daughter → Daughters inherit their X-Chr from their father in haplotype (allele association).

Y-Chr Transmission:

From Father to Sons: The Y-Chr is inherited in haplotype only from the father to his sons. Daughters do not inherit the Y-Chr.

The inheritance of gonosomal STRs is only one aspect of genetic analysis; they support other genetic markers, including autosomal STRs (aSTRs), to comprehensively understand an individual's genetic profile.

Y-Chr polymorphisms belong to the gonosomal STRs located on the sex chromosomes and, nowadays, are a crucial component of many forensic DNA investigations. Unlike aSTRs, gonosomal STRs are specific to the sex chromosomes and can be used to study genetic variations and inheritance patterns related to sex-linked traits.

Considering the topic of this thesis, it is appropriate to provide further information on the Y-Chr. The pseudo-autosomal regions PAR1 and PAR2, located at the

extremes of the short and long arms of the Y-Chr, are the only regions of X and Y-Chr involved in crossing over during meiosis.

The remaining portion of the Y-Chr between PAR1 and PAR2, approximately 95% of the Y-Chr, is known as the non-recombining region of the Y-Chr (NRY) and better defined as the male-specific region (MSY). It is exclusively transmitted to male offspring without genetic recombination, i.e. in haplotype (lineage markers). The MSY, in addition to being the site where mutations responsible for genetic variability accumulate during meiosis, is characterized by the presence of 8 palindromic regions (P1-P8), one of which is on the short arm and the remaining seven on the long arm, sites of intra-chromosomal gene conversion and in which Y-STRs are present. The mutation rate of Y-STRs is similar to that of autosomal, and it is estimated to be in the range of 10^{-3} .

Y-STRs are single-copy loci that generate only one allelic variant (amplicon) per marker analyzed following the PCR amplification process, providing an overall profile in hemizyosity. However, there are some exceptions: some Y-STRs, defined as multi-copy markers, originate an allelic pattern characterized by two amplification products as they are located in direct duplicated or inverted portions of the Y chromosome generated following gene conversion to palindromic regions. This is the case, for example, of DYS385 a/b, present in P4 on the long arm of the Y-Chr, in two regions spaced about 40.000 bp apart, whose single set of primers generates two amplicons referred to generically as *a* and *b*, which may be the same or different in the population. Another multi-copy Y-STR is the DYF387S1.

Inherited exclusively paternally in the haplotype, they make it possible to attribute a subject's appointment to a specific male genealogical tree but not to discriminate the grade of parenthood within it.

1.1 Brief History

The first genetic polymorphism in the human Y-Chr was identified in 1985 [Casanova et al., 1985; Lucotte & Ngo, 1985]. It took approximately seven years before Roewer et al. (1992) identified the first Y-STR [Roewer et al., 1992]. Subsequently, Roewer and Epplen applied this discovery to the first Y-STR analysis forensic case [Roewer & Epplen, 1992].

Between 1996 and 1997, two pivotal developments in Y-Chr research occurred. First, the establishment of the *minimal set* of Y-STR markers, consisting of eight loci on the Y-Chr (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS385 a/b and DYS393) demonstrated high polymorphism levels in two closely related human populations (Dutch and German) [Roewer et al., 1996].

Second, the first mutation rate study of these loci was published [Heyer et al., 1997]. It would wait until 2003 to have the first commercial multiplex for the amplification of these markers on Y-Chr, limited to only five/six loci [S K Sinha, 2003], and 2005 for the introduction of a larger commercial multiplex (12 Y-STRs) produced by Promega (PowerPlex Y System) [Krenke et al., 2005].

From 2005 to the present, a Literature review on PubMed reveals that approximately 2.000 papers related to the Y-Chr and its polymorphisms have been published, averaging 114 publications per year. The number of publications has sharply declined recently, with only 67 papers published in 2023. This decrease can be attributed to several factors, including the growing adoption of Next-Generation Sequencing (NGS) technologies in forensic genetics laboratories worldwide and challenges associated with deconvolution of male DNA mixtures, which have hindered the consistent development of new innovations in Y-STR research.

For a more detailed timeline of developments, please refer to the work of Peter de Knijff [De Knijff, 2022].

1.2 Y-Chr polymorphisms

Two types of Y-Chr polymorphisms are commonly employed in forensic analyses, either independently or in combination: Y-STRs and Y-Single Nucleotide Polymorphisms (SNPs).

The most widely used commercial Y-Chr amplification kits, including the *PowerPlex® Y23 System* (Promega), *Yfiler™ Plus PCR Amplification Kit* (Applied Biosystems), *Investigator Argus Y-28 QS Kit* (Qiagen), and *ForenSeq DNA Signature Prep Kit* (Verogen), target an average of 24 Y-STR loci. Among the seventeen Y-STR common to all four amplification kits, two are trinucleotide (DYS392 and DYS481), one is pentanucleotide (DYS438), one is hexanucleotide (DYS448), while the remaining loci are tetranucleotide repeats (DYS391, DYS389

I, Y-GATA-H4, DYS437, DYS533, DYS19, DYS389 II, DYS439, DYS390, DYS635, DYS576, DYS570, and DYS385 a/b).

The 13 rapidly mutating Y-STR loci (RM Y-STRs) identified by Ballantyne et al. in 2010 are particularly relevant in forensic applications. These include five multicopy systems (DYF387S1, DYF404S1, and DYS526 with two copies; DYF399S1 with three copies; and DYF403S1 with four copies) and eight single-copy Y-STR loci (DYS449, DYS518, DYS547, DYS570, DYS576, DYS612, DYS626, and DYS627) [Ballantyne et al., 2010].

Closely related male relatives can be distinguished by RM Y-STR along paternal lineages, providing a distinct advantage in specific forensic contexts.

In contrast, the development of Y-SNP multiplex systems for routine forensic use has lagged behind that of Y-STRs despite evidence that large numbers of Y-SNPs can be effectively genotyped [Brión, 2005; Ralf et al., 2019; De Knijff, 2022].

Y-SNPs, which are slow-evolving bi-allelic markers, are primarily used in reconstructing human phylogenetic trees, providing insights into ancestral origins and facilitating the examination of evolutionary migration patterns [Claerhout et al., 2021; Qian et al., 2017].

To ensure clarity in terminology, it is essential to distinguish between *haplotype*, *haplogroup*, and *lineage*. As defined by De Knijff in 2000, a haplotype refers to “Distinct Y chromosomes identified by STRs”, a haplogroup describes “Distinct Y chromosomes, defined solely on the basis of unique mutation event (UME) character states”, and a lineage represents the combination of both [De Knijff, 2000].

When two male subjects share the same haplotype, they are not necessarily related.

Three scenarios can explain this phenomenon:

- a) Chance: two males can coincidentally share indistinguishable Y-haplotypes on a similar or a distinct Y-haplogroup background, sharing these Y-haplotypes by state (IBS);
- b) Descendants: Numerous earlier pedigree studies with a deep-rooted focus have illustrated that men who trace back to a common paternal ancestor

within the last 20 generations may exhibit a matching Y-STR haplotype through descent (IBD);

c) Repeat homoplasy.

1.3 Y-Chr polymorphisms forensic applications

Y-STR profiling in forensic genetics involving male individuals provides critical supplementary information to autosomal STR analysis and can be decisive in some instances.

Y-STRs are applicable in various forensic contexts:

- Crime-scene samples with male/female mixtures → Y-STRs are particularly useful in cases with male/female DNA mixtures because they do not suffer from preferential amplification. This allows for the isolation of the male DNA profile even in mixtures with a high female-to-male ratio (e.g., 1:3000). This scenario is common in sexual assault investigations, where Y-STRs help differentiate the female genetic material (e.g., vaginal epithelium from the victim) from male genetic material (e.g., spermatozoa, seminal fluid, or epithelial cells from the perpetrator) [K Honda et al., 1999; Dekairelle & Hoste, 2001; Roewer, 2009].
- Paternity testing and familial searching → In deficiency cases where the alleged biological father's genetic profile is unavailable, Y-STR analysis can be used to infer the paternal lineage by analyzing the Y-STR profiles of male relatives (e.g., sons, brothers, grandfathers, uncles, cousins) [Roewer & Epplen, 1992; De Knijff et al., 1997; Kayser et al., 1997]. This approach also applies to familial searching, mass disaster victim identification, and missing person investigations [Kayser, 2017; Mo et al., 2019].
- Establishing relationships between two or more males → Y-STR profiling can be used to trace relationships between males, particularly by employing a set of 13 rapidly mutating Y-STR (RM Y-STR) loci, which enhance the discriminatory power between related and unrelated male profiles [Forouzesh et al., 2022]. Analyzing RM Y-STRs is

particularly useful for identifying more distant male relatives based on mutations and focusing on closely related individuals by excluding those lacking the relevant mutations. This strategy is valuable in forensic searches for an unknown male perpetrator whose DNA is not in reference databases [Kayser, 2017].

- Male DNA dragnets → In cases involving male perpetrators, solving the crime may require mass screening through DNA collection, where large numbers of individuals voluntarily submit biological samples for comparison with the DNA profile obtained from the crime scene. Since it is assumed that the actual perpetrator is unlikely to participate in the DNA screening, the use of Y-STR analysis in this context focuses on identifying male relatives of the perpetrator. This strategy can help narrow the investigation to a specific family group, aiding in identifying the perpetrator [Kayser, 2017].
- Paternal bio-geographic ancestry → Y-STRs also provide insight into an individual's paternal bio-geographic origins, revealing the geographic region from which their paternal ancestors descended. This information can be instrumental in guiding investigators in cases involving unknown male perpetrators, as well as in missing persons investigations and disaster victim identification, mainly when no prior information about the individual is available [Kayser, 2017; Ambrosio et al., 2020].

REFERENCES

- Ambrosio, I. B., Braganholi, D. F., Orlando, L. B. M., Andrekenas, N. C., da Mota Pontes, I., da Silva, D. A., Astolfi-Filho, S., de Carvalho, E. F., Cicarelli, R. M. B., & Gusmão, L. (2020). Mutational data and population profiling of 23 Y-STRs in three Brazilian populations. *Forensic Science International: Genetics*, *48*, 102348. <https://doi.org/10.1016/J.FSIGEN.2020.102348>
- Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., Van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., Von Wurmb-Schwark, N., De Knijff, P., Labuda, D., Ne Vézina, H., Knoblauch, H., Lessig, R., Roewer, L., ... Kayser, M. (2010). *ARTICLE Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications*. <https://doi.org/10.1016/j.ajhg.2010.08.006>
- Brión, M. (2005). Y chromosome SNP analysis using the single-base extension: a hierarchical multiplex design. *Methods in Molecular Biology (Clifton, N.J.)*, *297*, 229–242. <https://doi.org/10.1385/1-59259-867-6:229/COVER>
- Casanova, M., Leroy, P., Boucekkine, C., Weissenuach, J., Bishop, C., Fellous, M., Purrello, M., Fior, G., & Siniscalco, M. (1985). A human Y-Linked DNA Polymorphism and Its Potential for Estimating Genetic and Evolutionary Distance. *Science*, *230*(4732), 1403–1406. <https://doi.org/10.1126/SCIENCE.2999986>
- Claerhout, S., Verstraete, P., Warnez, L., Vanpaemel, S., Larmuseau, M., & Decorte, R. (2021). CSYseq: The first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-STRs to unravel worldwide human population genetics. *PLoS Genetics*, *17*(9). <https://doi.org/10.1371/journal.pgen.1009758>
- Dekairelle, A. F., & Hoste, B. (2001). Application of a Y-STR-pentaplex PCR (DYS19, DYS389I and II, DYS390 and DYS393) to sexual assault cases. *Forensic science international*, *118*(2-3), 122–125. [https://doi.org/10.1016/s0379-0738\(00\)00481-3](https://doi.org/10.1016/s0379-0738(00)00481-3)
- De Knijff, P. (2000). INVITED EDITORIAL Messages through Bottlenecks: On the Combined Use of Slow and Fast Evolving Polymorphic Markers on the Human Y Chromosome. In *Am. J. Hum. Genet* (Vol. 67).
- De Knijff, P. (2022). On the Forensic Use of Y-Chromosome Polymorphisms. In *Genes* (Vol. 13, Issue 5). MDPI. <https://doi.org/10.3390/genes13050898>
- De Knijff, P., Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., ... Roewer, L. (1997). Chromosome Y microsatellites: Population genetic and evolutionary aspects. *International Journal of Legal Medicine*, *110*(3), 134–140. <https://doi.org/10.1007/S004140050052/METRICS>
- Forouzesh, M., Irani, S., Soleimani, A., & Monabati, S. J. (2022). Application of Y-STR, DIP-STR and SNP-STR Markers in Interpretation of Forensic Genetic Profiling: A Narrative Review. *Iran J Public Health*, *51*(7), 1538–1545. <https://creativecommons.org/licenses/by-nc/4.0/>

- Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E., & De Knijff, P. (1997). Estimating Y Chromosome Specific Microsatellite Mutation Frequencies using Deep Rooting Pedigrees. *Human Molecular Genetics*, 6(5), 799–803. <https://doi.org/10.1093/HMG/6.5.799>
- Kayser, M. (2017). Forensic use of Y-chromosome DNA: a general overview. In *Human Genetics* (Vol. 136, Issue 5, pp. 621–635). Springer Verlag. <https://doi.org/10.1007/s00439-017-1776-9>
- K Honda, L Roewer, & P de Knijff. (1999). Male DNA typing from 25-year-old vaginal swabs using Y chromosomal STR polymorphisms in a retrial request case. *J Forensic Sci.*, 44(4), 868–872.
- Krenke, B. E., Viculis, L., Richard, M. L., Prinz, M., Milne, S. C., Ladd, C., Gross, A. M., Gornall, T., Frappier, J. R. H., Eisenberg, A. J., Barna, C., Aranda, X. G., Adamowicz, M. S., & Budowle, B. (2005). “Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex” [Forensic Sci. Int. 148 (1) (2005) 1–14]. *Forensic Science International*, 151(1), 111–124. <https://doi.org/10.1016/J.FORSCIINT.2005.02.008>
- Lucotte, G., & Ngo, N. Y. (1985). p49f, A highly polymorphic probe, that detects TaqI RFLPs on the human Y chromosome. *Nucleic Acids Research*, 13(22), 8285. <https://doi.org/10.1093/NAR/13.22.8285>
- Mo, X. T., Zhang, J., Ma, W. H., Bai, X., Li, W. S., Zhao, X. C., & Ye, J. (2019). Developmental validation of the DNATyper™Y26 PCR amplification kit: An enhanced Y-STR multiplex for familial searching. *Forensic Science International: Genetics*, 38, 113–120. <https://doi.org/10.1016/J.FSIGEN.2018.10.008>
- Qian, X., Hou, J., Wang, Z., Ye, Y., Lang, M., Gao, T., Liu, J., & Hou, Y. (2017). *Next Generation Sequencing Plus (NGS+) with Y-chromosomal Markers for Forensic Pedigree Searches*. <https://doi.org/10.1038/s41598-017-11955-x>
- Ralf, A., van Oven, M., Montiel González, D., de Knijff, P., van der Beek, K., Wootton, S., Lagacé, R., & Kayser, M. (2019). Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing. *Forensic Science International: Genetics*, 41, 93–106. <https://doi.org/10.1016/J.FSIGEN.2019.04.001>
- Roewer, L. (2009). Y chromosome STR typing in crime casework. In *Forensic Science, Medicine, and Pathology* (Vol. 5, Issue 2, pp. 77–84). <https://doi.org/10.1007/s12024-009-9089-5>
- Roewer, L., Arnemann, J., Spurr, N. K., Grzeschik, K.-H., & Epplen, J. T. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts*. In *Hum Genet* (Vol. 89). Springer-Verlag.
- Roewer, L., & Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Science International*, 53(2), 163–171. [https://doi.org/10.1016/0379-0738\(92\)90193-Z](https://doi.org/10.1016/0379-0738(92)90193-Z)
- Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., & De Knijff, P. (1996). Analysis of Molecular Variance (Ampva) of Y-Chromosome-Specific

Microsatellites in Two Closely Related Human Populations. *Human Molecular Genetics*, 5(7), 1029–1033. <https://doi.org/10.1093/HMG/5.7.1029>

S K Sinha. (2003). Forensic Casework Applications Using Y-PLEX™ 6 and Y-PLEX™ 5 Systems. *Forensic Sci Rev.*, 15(2), 199–203.

CHAPTER 2

DNA MIXTURES	22
2.1 DNA mixtures interpretation (aSTRs)	25
2.1.1 Determining the number of contributors	30
2.1.2 Frequentist approaches	30
2.1.2.1 Random Match Probability (RMP)	30
2.1.2.2 Random Man not Excluded (RMNE)	30
2.1.3 Likelihood Ratio approaches	32
2.1.3.1 Semi-continuous model.....	33
2.1.3.2 Continuous model	35
2.2 DNA mixtures interpretation (Y-STRs)	37
2.2.1 Y chromosome is inherited in allelic association	37
2.2.2 Lack of a unique model for calculating haplotype frequencies	38
2.2.3 Assembling pairs, triplets, and quadruplets as haplotypes	40
2.2.4 Not all conceivable haplotypes are real	41
2.2.5 No developed strategies	41
2.2.6 Low computing power of personal computers	41
REFERENCES	43

DNA MIXTURES

DNA mixture is a typical forensic evidence where DNA from two or more individuals is present in a single sample, in different proportions. This happens when two or more individuals concur to create a trace at a given crime scene.

However, a mixture can also result from the contamination of an evidentiary sample or the intermingling of two foreign (i.e., non-evidentiary) DNA sources.

Contamination can occur at the crime scene, thereabouts, or laboratory bench levels.

The enhanced sensitivity of commercial kits for amplifying forensic genetic markers has significantly increased the detection of mixed DNA profiles in molecular investigations conducted for criminal identification.

As with single-source DNA traces, mixed DNA profiles must be carefully evaluated to determine whether they originated from activities directly associated with the commission of the crime (*activity level analysis*).

The molecular analysis of mixed DNA traces aims to identify the individuals contributing to the DNA and clarify their roles in criminal acts. For instance, if male and female DNA are found intermingled in a vaginal swab, this can provide crucial information about the involvement of both individuals in the context of a sexual assault, shedding light on the nature of their participation, whether as perpetrator or victim.

Once the trace is typed and a mixed genotype is assigned to the item, the DNA profile of a Person of Interest (POI) can be compared to the mixed evidence. A positive association is established if all the alleles carried by the POI are also present in the evidence sample.

When a trace is supposed to contain a DNA mixture, the forensic geneticist will be faced with at least two preliminary issues:

- a) determining the number of contributors, and
- b) assigning distinct genotype sets to each contributor.

The number of detected allele peaks in the DNA profile typically provides valuable clues for addressing both issues.

To simplify with an elementary example: in DNA mixtures involving two contributors, each locus can display one, two, three, or four distinct alleles. This variability depends on whether both individuals share identical alleles at a locus or if they contribute different alleles. (Figure 1) [Butler & John M, 2015].

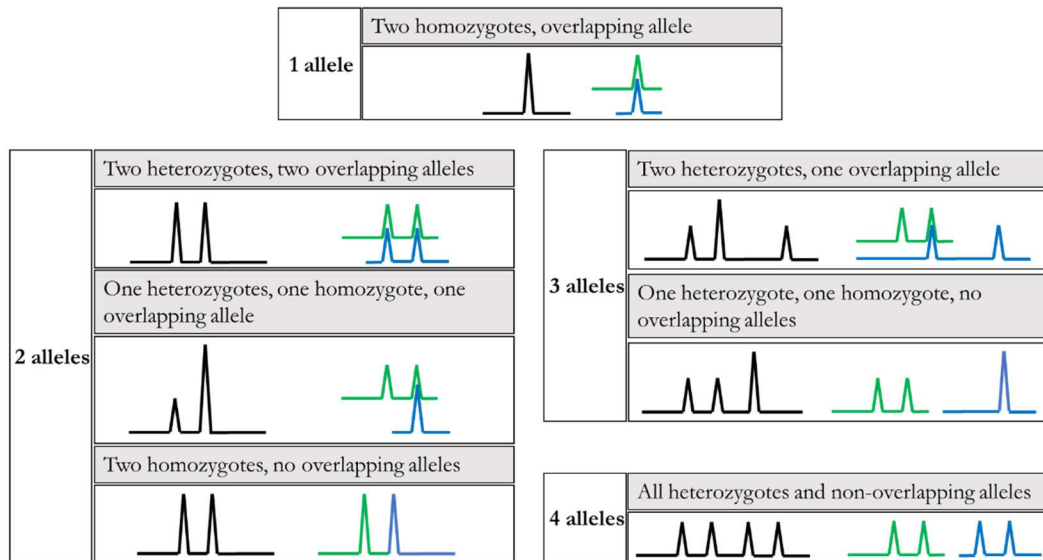


Fig. 1 DNA mixtures involving two contributors. The observed profile indicates variations in peak heights, reflecting the effects of allele stacking attributable to shared alleles and a 1:1 mixture ratio.

Under this evidentiary perspective (or ‘hypothesis’), several *thumb-rule* have been issued to solve both issues quickly. For example, the ‘German Stain Commission’ devised a threefold classification (Figure 2) [Schneider et al., 2009]:

- *Type A*: no major contributor is detectable;
- *Type B*: major and minor contributors are distinguishable;
- *Type C*, low-level DNA can produce stochastic effects.

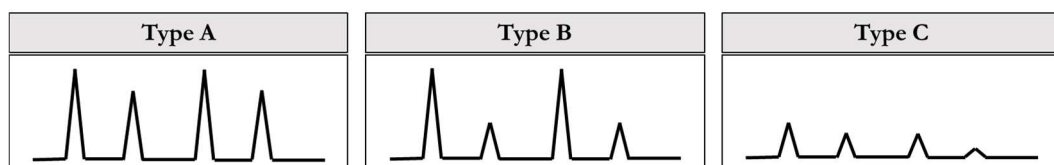


Fig. 2 Graphical representation of DNA mixture classification given by the Germain Stain Commission.

The number of alleles at each locus and their proportions are helpful but not definitive indicators when interpreting DNA mixtures. As a general guideline, a maximum of four alleles at a locus suggests a mixture of two heterozygotes, while five or six alleles indicate three or more contributors. However, DNA mixtures can be far more complex than suggested by thumb-rules.

First, even with only four alleles at a locus, a mixture may involve more than two contributors, depending on the circumstances. Four peaks represent the maximum genetic heterogeneity possible for two contributors, typically when their genetic backgrounds are highly divergent, such as when they belong to populations with significantly different allele frequencies. On the other hand, when contributors are genetically similar (e.g., close relatives), it is more likely to observe only two or even a single peak at a locus due to shared alleles.

Second, some contributors may provide a suboptimal amount of DNA, leading to lower signal detection in the analysis.

Third, the DNA of certain contributors may degrade, causing alleles to drop-out during the amplification process. This degradation, influenced by random "stochastic" effects, can result in the loss of some alleles initially present in the evidence, a phenomenon known as *allele drop-out*. This complicates the interpretation of mixed profiles and necessitates caution in drawing conclusions [Bieber et al., 2016].

DNA mixtures underlie various and disparate scenarios: a sexual assault, a close-quarter murder (where, for example, two types of blood happen to mix), a criminal act of any kind involving effraction (with various cells superimposing at a doorknob), a close-quarter fight and the victim's fingernail clippings taken during the autopsy; etcetera [Bieber et al., 2016].

Contamination risks exist at crime scenes or in the laboratory, which implies the need to gather a collection of reference samples from the laboratory staff.

PCR reagents and plastic disposables may also contain low-level contaminating DNA, emphasizing the importance of manufacturer reference samples to identify and address potential sources of contamination during DNA testing.

All this adds complexity to the case of unravelling the mixture.

2.1 DNA mixtures interpretation (aSTRs)

Interpretation of mixtures “continues to be the most difficult challenge that faces scientists, lawyers, and judges” [Gill, 2019].

Interpreting DNA mixtures involves two main tasks: (1) deducing potential genotypes that could represent the individuals contributing to the sample (a process often called *deconvolution* of the mixture components) and (2) determining the strength of the evidence that supports the inclusion of a POI (person of interest or candidate to the role of mixture contributors) in the DNA profile found in the evidence.

A DNA mixture can be interpreted following the steps represented in Figure 3. In brief, once the test element is collected, the DNA is extracted, quantified and amplified by polymerase chain reaction (PCR) using a predefined set of genetic markers of STR loci. The obtained amplicons are then separated and detected by capillary electrophoresis. The results are then interpreted by the operator, compared with the profiles of the reference samples and statistically evaluated. This process can be divided into two stages: (1) *measurement*, which reflects the physical properties of the sample and includes all the necessary preparations for the generation of a DNA profile, graphically represented in an electropherogram (EPG), and (2) the *interpretation* of the genetic profile, which culminates in the preparation of a written report, including statistical analysis that assigns evidentiary weight to the analyzed evidence. The interpretation phase depends on the DNA analyst attributing values not inherently part of the sample [Butler et al., 2021; Gill, 2019].

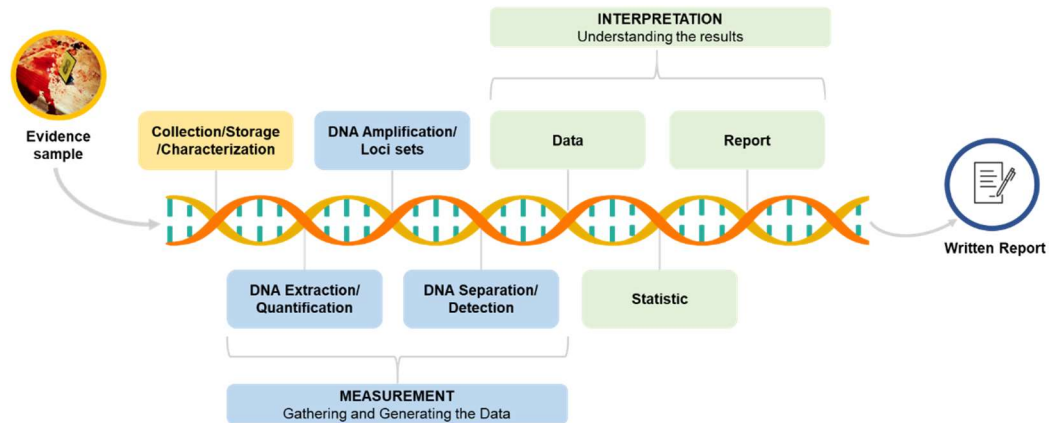


Fig. 3 Steps for a correct interpretation of a DNA mixture. Figure inspired by Butler et al., 2021.

In 2006, nine fundamental recommendations for interpreting DNA mixtures and the steps that must be followed were published by the DNA Commission of ISFG [Gill et al., 2006]. Among the recommendations, the most important points are the following [Butler et al., 2021]:

- i. As peak heights decrease, the variability in relative peak heights increases due to stochastic variation during the PCR amplification of the mixture components;
- ii. If sufficient DNA is available, the reproducibility of mixture results and the consistency of relative peak heights among mixture components should be confirmed through repeated testing;
- iii. Reducing the number of simultaneously amplified loci enhances the balance of relative peak heights.

The seven steps proposed by the DNA Commission of ISFG are represented in Figure 4:

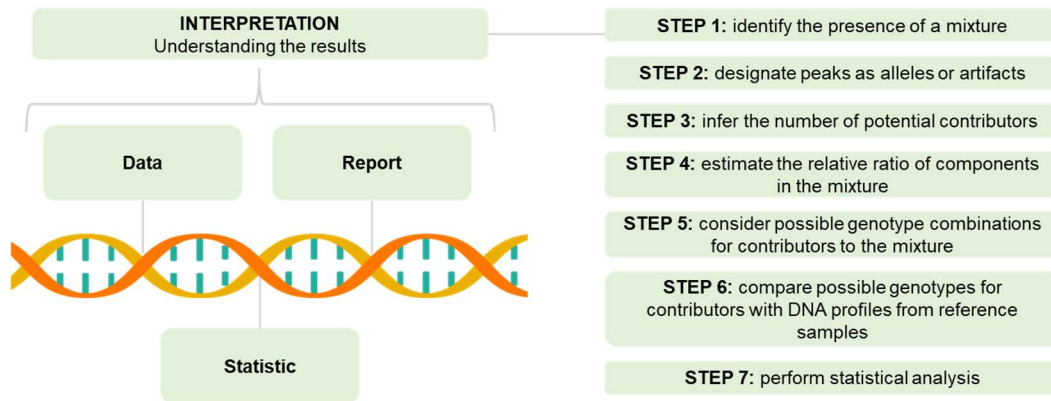


Fig. 4 Steps for correctly interpreting a DNA mixture proposed by the DNA Commission of ISFG. Figure inspired by Butler et al., 2021.

There can be autosomal and haploid profiles, multiallelic profiles, and SNP mixtures (autosomal or haploid). So far, forensic experts have preferentially addressed autosomal multiallelic mixtures, with little attention paid to the haploid ones and the SNPs.

At present, in the matter of mixed aSTRs, two main models of interpretation have been devised:

- i. Pure deconvolution of a mixture: which is equivalent to saying that the expert tries to determine the whole sets of possible combinations to be elicited from the evidence;
- ii. A POI-to-evidence comparison: what matters in this case is answering the only question of whether the POI is present in the evidence, irrespective of other possible issues (including the number of possible contributors).

Several factors contribute to increased complexity in connection with these two interpretation models:

- i. Determining the mixture ratio. When the alleles of a DNA mixture, due to the low amount of DNA of the minor contributor, are highly unbalanced, stochastic amplification effects (such as drop-out and drop-in) may arise due to random sampling variations. It follows that the true alleles of the minor

contributor may either not be genotyped at all or not be correctly interpreted if mistaken for stutters of the major contributor alleles.

- ii. Determining the appropriate number of contributors in complex mixtures containing very many in low amounts. Complex DNA mixtures tend to contain a low amount of DNA from each contributor, and consequently, drop-outs are more frequent. Drop-in phenomena (erratic alleles that have nothing to do with legitimate evidence) also complicate things. In such complex mixture cases, it becomes increasingly difficult to decipher complete genotypes [Yang et al., 2019].

Jan Evett I.W., Bruce Weir and John Buckleton pioneered the POI-to-Evidence model of mixture interpretation [Evett et al., 1991; Weir et al., 1997; Evett & Weir, 1998]. Mark Perlin was the first to address the issue of mixture deconvolution [Perlin & Sinelnikov, 2009].

Formerly, a rough early version of mixture interpretation was drawn from the POI-to-single source trace case.

In cases involving single-source samples or when an individual's profile can be determined from a DNA mixture, the *Random Match Probability* (RMP) method is applicable, representing the probability of matching the POI profile to a person chosen randomly in the population. This rather rough method was applied to mixtures, too, by noticing a POI-to-trace compatibility (based on the only assumption that all POI alleles have appeared in the complex evidence of the trace). This statistic assumes value 1 to the numerator and the genotype frequency at the denominator and consequently depends only on the allele frequencies drawn from the relevant populations. A low RMP suggests that finding that particular DNA profile in an unrelated, randomly selected individual will be rare.

In some mixtures with a big difference between a major contributor and the rest of the trace, the RMP method can be adopted to approximate the probability of that profile being in the trace.

However, more often, the quantitative ratio among contributors does not allow this method to be conveniently applied, and a full deconvolution of all combinations is preferable. When no person of interest is available, this leads to a deconvolution,

which is the equivalent of listing all combinations to be envisioned by any given example of evidence.

So, for example:

$E(A; B)_{\text{autosomal}} ; \text{two contributors} \rightarrow (AA; BB); (AA; AB); (AB; AB); (AB; BB)$

$E(A; B; C)_{\text{autosomal}} ; \text{two contributors} \rightarrow (AA; BC); (BB; AC); (CC; AB); (AB; AC); (AB; BC); (AC; BC)$

$E(A; B; C; D)_{\text{autosomal}} ; \text{two contributors} \rightarrow (AB; CD); (AC; BD); (AD; BC)$

If reference genotypes from suspects or victims are available, the number of combinations is restricted to a few.

For example:

$E(A; B; C)_{\text{autosomal}} ; \text{two contributors}; \text{POI} = AB \rightarrow (AB; BC), (AB; AC), \text{ or } (AB; CC)$

In that case, the prosecutor or the defence may diverge as to the only presence of the POI in the evidence. Then, a Likelihood Ratio can be envisioned, as in the following example:

Prosecutor (AB is certain) $\rightarrow (AB; BC), (AB; AC), \text{ or } (AB; CC)$

Defence (no certainty on the POI) $\rightarrow (AA; BC); (BB; AC); (CC; AB); (AB; AC); (AB; BC); (AC; BC)$

$$LR = \frac{P(AB; BC) + P(AB; AC) + P(AB; CC)}{P(AA; BC) + P(BB; AC) + P(CC; AB) + P(AB; AC) + P(AB; BC) + P(AC; BC)}$$

P represents the genotype frequency of every relevant genotype mentioned in the numerator or the denominator.

This is tantamount to determining the WoE conditional to the hypotheses put forward by the analyst.

We will systematically develop both issues in the following chapters.

2.1.1 Determining the number of contributors

The number of contributors can never be defined with certainty. As a consequence, this parameter is more often an arbitrary assumption. One should, at large, say that any genetic evidence can account for whatever number of contributors.

In practice, bonding the evidence to a reasonable/minimum number of contributors is the essential starting point for interpretation.

This is often justified by assuming thumb rules. If a locus has n distinct alleles, the minimum number of contributors is defined by $n/2$ since each contributor carries a maximum of two different alleles per locus.

Thumb rules are mostly a naïve way to address the issue. Lauritzen et al. and Egeland et al. have introduced a likelihood-based estimator of contributors [Lauritzen & Mortera, 2002; Egeland et al., 2003].

2.1.2 Frequentist approaches

2.1.2.1 Random Match Probability (RMP)

As previously described, if there are significant differences in peak heights within a mixture, the major and minor components can be easily separated into individual profiles. In this case, it is possible to calculate the *Random Match Probability* (RMP), also known as *Coincidence Probability* (CI), for each profile as if the mixture originated from a single source. This application can also be extended to profiles obtained after deconvoluting the DNA mixture, a process referred to as *Modified RMP* (mRMP). However, in this scenario, the forensic expert can accurately and with minimal ambiguity interpret only the profile of the major contributor, as the minor contributor may experience allele masking or potential allele drop-out. Therefore, this approach is suitable only when ‘unambiguous’ genotypes can be deduced from peak height information and not in cases of low DNA quantities, where the chance of encountering stochastic effects increases [T. Bille et al., 2013; SWGDAM, 2017; Butler & John M, 2015]

2.1.2.2 Random Man not Excluded (RMNE)

When peak heights in a mixture are the same in quantity, the DNA becomes a conundrum. A way to escape the interpretation of such complex evidence is to

calculate the probability of exclusion from the evidence profile through a binary approach named *Random Man Not Excluded* (RMNE).

It is considered a population genetics measure used to assess the likelihood that an individual randomly selected from the general population would not be excluded from matching a specific genetic data set.

According to this approach, all possible genotype combinations have the same weight; in fact, it uses *Combined Probability of Inclusion* (CPI) statistics that involve the sum of all observed alleles at each locus and their subsequent product to obtain the overall profile inclusion probability [Butler & John M, 2015]. In other words, the CPI defines the probability that a randomly chosen individual would be included in the mixture as a possible contributor.

From a mathematical point of view, given the DNA evidence with alleles A_1, A_2, \dots, A_n in a locus l , and their frequencies f_1, f_2, \dots, f_n , the RMNE is calculated as follows [Haned, 2010]:

$$RMNE = 1 - \left(\sum_{i=1}^n f_i \right)^2$$

The main limitation of this statistical method is its inefficiency in the presence of *low-template DNA* (LT-DNA) samples [Bille et al., 2014], caused by not considering the number of contributors to the sample, peak heights, and the presence/absence of drop-outs or drop-ins. In fact, following the ISFG recommendations, the RMNE approach applies only to unambiguous DNA profiles. In cases where the crime stain profile has low-level DNA and specific minor alleles coincide in size with stutters of major alleles, and/or there is a possibility of drop-out, the RMNE method may not be conservative [Gill et al., 2006].

The frequentist approaches are significantly limited, mainly because they consider the probability of the evidence based on a single hypothesis (usually, a person taken randomly within the population, other than the suspect, contributes to the DNA mixture).

Essentially, a thorough evaluation of DNA evidence requires the consideration of at least one additional alternative hypothesis for the source of the DNA evidence [Haned, 2010]. Moreover, the inefficiency of binary methods in the presence of LT-DNA and the consequent inability to consider amplification replicates has prompted forensic experts to increasingly develop and exploit LR approaches.

2.1.3 Likelihood Ratio approaches

Likelihood Ratio approaches are statistical tools used in forensic DNA analysis to compare two alternative and mutually exclusive hypotheses. Specifically, it assesses the probability of generating a DNA mixture if the POI is a contributor versus the probability if an unknown, unrelated individual is the contributor.

In other words, the LR assesses whether a POI contributes to the sample E by reporting a quantity associated with the WoE.

Generally, the *prosecution hypothesis* (H_p) argues that the POI contributed to the sample, while the *defence hypothesis* (H_d) posits an unknown individual as the source.

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

The numerator of the LR involves the probability of STR data given the H_p ; the denominator is the probability of the STR data given the H_d .

Different scenarios can be hypothesized to calculate the LR of a mixture:

- Probable presence of **two contributors** in the mixture:

$$LR = \frac{POI + unknown}{2 unknowns}$$

- Probable presence of **three contributors** in the mixture:

$$LR = \frac{POI + 2 unknown}{3 unknowns}$$

An LR greater than 1 supports the prosecution hypothesis, favouring that the POI is a contributor. A higher LR provides stronger support for this hypothesis. Conversely, an LR of less than 1 supports the defence hypothesis, favouring the notion that an unknown individual is a contributor, with a lower LR offering stronger support for this defence hypothesis [John S. Buckleton et al., 2016].

Following the recommendation expressed by Gill P. et al. [Gill et al., 2006], a valid method to calculate the LR of mixtures is to follow the procedure reported by [Evetts et al., 1991] and [Weir et al., 1997].

2.1.3.1 Semi-continuous model

The interpretative principle of the semi-continuous methods is qualitative, defined as such because the calculation of the LR only considers the qualitative information (presence/absence of alleles) of the electropherograms obtained from the typing of the mixed evidence.

Gill et al. referred to this approach as a *non-restricted combinatorial approach* because all genotypic combinations of the unknown contributors are considered without any limitations [Gill et al., 2006].

The absence of a few alleles in a context of compatibility between the POI and the mixed evidence is justified by employing the drop-out hypothesis; conversely, the presence of extra alleles is justified by utilizing the drop-in hypothesis.

The fact that qualitative models use little information for the LR calculations may lower their sensitivity in complex situations with more than two/three contributors [Costa et al., 2022].

According to the semi-continuous model, all the genotypes generated from evidence have the same probability of existing net of differences in genotype frequencies.

In the two-contributor's evidence, considering four alleles in a locus, three scenarios are possible (six if the reciprocal genotype combinations are considered), as illustrated in Figure 5.

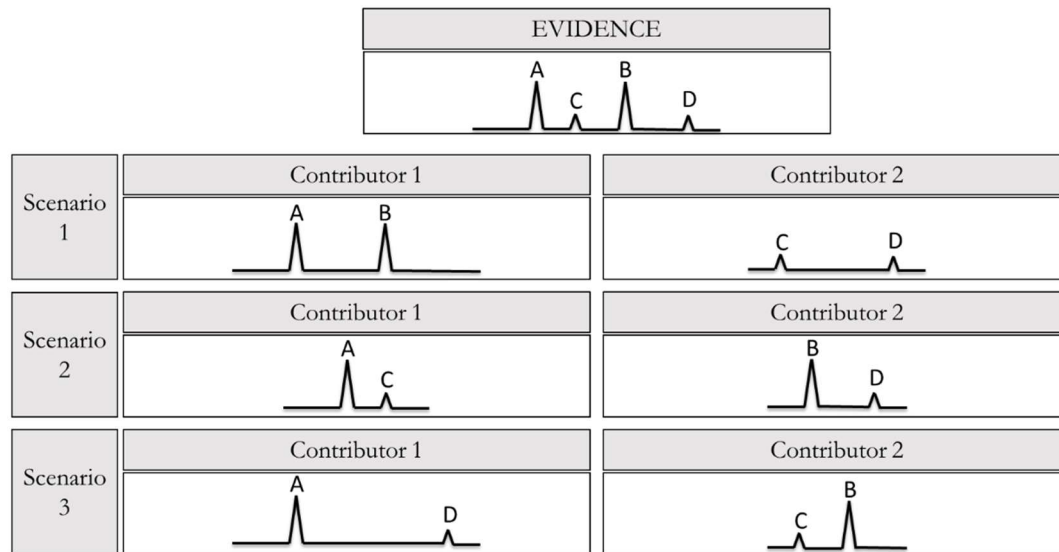


Fig. 5 Possible deconvolution scenarios of the two contributor's evidence analyzed with a semi-continuous model.

Informatics tools designed to analyze and interpret DNA results from forensic samples quantify genetic evidence by computing a probative value based on information extracted from electropherograms.

Semi-continuous software

In 2007, some authors embarked on developing probabilistic genotyping using qualitative software that considered allele drop-out and drop-in but did not model peak heights. The initial software, *LoComatioN*, was introduced by James Curran at the Forensic Science Service (UK) [Gill et al., 2007]. Hinda Haned later re-programmed the model, resulting in *LRmix*, written in *R* and included in the *forensim* package [Haned, 2011; Haned et al., 2015].

In 2013, the Netherlands Forensic Institute (NFI) adopted *LRmix*, rewriting it in Java and rebranding it as *LRmix Studio*. Widely embraced in Europe and beyond, NFI further enhanced *LRmix Studio* to introduce *SmartRank*, a more effective database search engine than the CODIS search engine [Benschop et al., 2017; Gill et al., 2021].

Currently, in Italian forensic genetics laboratories, the most widely used qualitative software for the deconvolution of DNA mixture is *LRmix Studio*, a cost-free, open-source expert system designed for interpreting forensic DNA profiles, especially those involving complex DNA mixtures. It facilitates the assessment of the

probative value of forensic DNA profiles based on autosomal STR markers. Notably, this model addresses uncertainty arising from allelic drop-out and drop-in phenomena in DNA profiles. The software estimates these quantities from the provided data and utilizes them to generate likelihood ratios. *LRmix studio* was programmed after the development of the likelihood ratio model by Haned H. and Gill P. [Haned et al., 2012; Gill & Haned, 2013].

2.1.3.2 Continuous model

In continuous model calculation of the LR, the quantitative information (peak height/area) is drawn from the electropherograms.

Adopting this interpretive perspective required a laborious change of perspective, which entailed giving up some long-standing empirical patterns.

Formerly, the average forensic geneticist would instinctively look after a major and some minor components in the evidence. A significant imbalance in peak height among contributors would lead to a straightforward interpretation; if not, the evidence was classified as ambiguous [Gill et al., 2006].

An original model of quantitative treatment was eventually described by Cowell et al. and extended to some well-known classes of electrophoretic artefacts [Cowell et al., 2013].

According to the continuous model, all genotype combinations have a definite probability of existing within the mixture.

So, if, for example, we refer to the 2 PM example illustrated in section 2.1.3.1 and if considering four alleles within a locus, scenario 1 is more probable than scenarios 2 and 3, as illustrated in Figure 6.

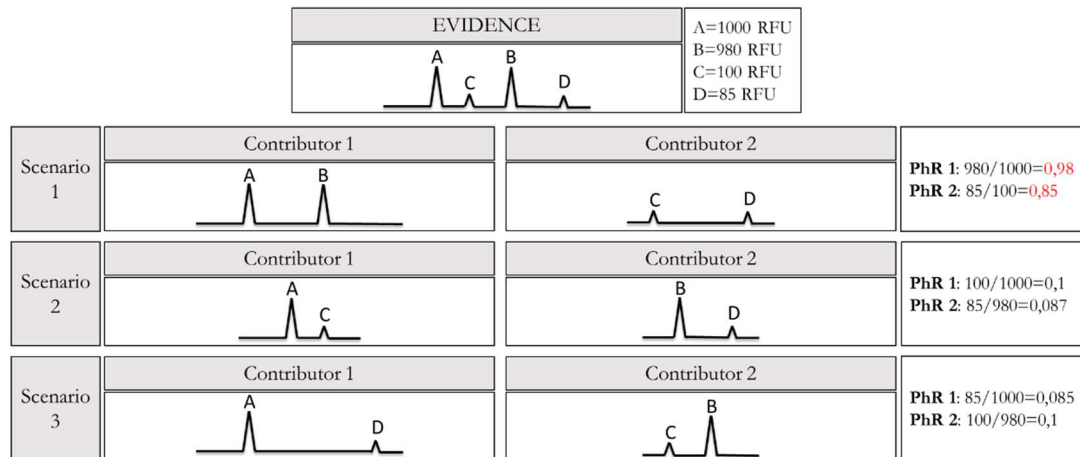


Fig.6 Possible deconvolution scenarios of the two contributor's evidence analyzed with a continuous model.

Continuous software

Evetts at first, and later Cowell introduced early models explaining variation in peak area observations, defining underlying normal and gamma distributions, respectively [Evetts et al., 1998; Cowell et al., 2007].

In 2013, Cowell, Graverson, and colleagues developed DNA mixtures interpretation models [Cowell et al., 2007; Graverson & Lauritzen, 2013] based on the gamma model and written in *R code*, requiring *HUGIN* for execution. Supported by the EuroForGen-Network-of-Excellence, Øyvind Bleka rewrote the gamma model as *EuroForMix* in *R* and *C++*. This program surpassed *DNAmixtures* with enhanced capabilities, including degradation parameterization and "theta-correction" (Fst). *EuroForMix* evolved into *DNAmatch2*, a database search tool incorporating the *forensim LRmix* module for searches in large national DNA databases. These modules were integrated into the user-friendly expert system, *CaseSolver*, for complex case analysis, offering visualization, automated comparison, deconvolution, weight-of-evidence evaluation, and reporting.

In 2019, NFI implemented *DNAMixStatX*, a statistical module based on *EuroForMix*, usable as a stand-alone application or within *DNAXs*, a comprehensive software suite developed by NFI for managing and interpreting DNA profiles in forensic casework since 2017. *DNAXs* undergoes continuous development to enhance functionality, improve DNA casework processes, and broaden its application scope.

Among the most widely used software in forensic genetics laboratories for the quantitative interpretation of mixed traces typed with autosomal markers are *EuroForMix* and *STRmix* [Taylor et al., 2013; Bleka et al., 2022].

Euroformix is a software tool designed to analyze DNA data. It includes procedures for maximizing likelihood functions using frequentist approaches and integrating them using Bayesian methods. The tool employs a gamma peak height model for both single and replicated data of various DNA types, such as STR, SNP, and MPS. Users can specify general hypotheses and conduct sensitivity analyses of unknown parameters using the Markov Chain Monte Carlo method. Additionally, *Euroformix* features procedures for deconvolution and database search and addresses issues like stutters, allele drop-out, and allele drop-in during the analysis.

2.2 DNA mixtures interpretation (Y-STRs)

A general framework for calculating likelihood in the case of autosomal DNA mixtures has been described above (Chapter 2.1). Still, the diploid approach cannot be applied directly to the Y-Chr, which is haploid.

Surprisingly, the relevant Literature lacks the principles and methodologies applied to Y-Chr mixtures.

Here, several significant problems make the theoretical elaboration and technical progress difficult.

2.2.1 Y chromosome is inherited in allelic association

The Y chromosome is inherited as a big haplotype because of the lack of recombination.

Every individual has a haplotype inherited as a whole from his male lineage.

The presence of haplotypes is a serious obstacle to the elaboration of comprehensive tables of frequency data to adopt in the procedures of individual identification. Specifically, the frequency of every haplotype is a property of the male lineage, not of the individuals aligned with it; also, sampling independent lineages is much more complex than sampling unrelated individuals.

The probability of selecting two male-related individuals in a random survey of a large cosmopolitan population may be small. The same cannot be said of small/

isolated/rural populations. Unfortunately, the issue of random sampling has no straightforward solution. Gathering as much information as possible about the number of male lineages to exist in a population and making a census of them may be an elusive task.

Expanding the number of loci entering a haplotype may involve re-sampling the population if the original DNAs on which the previous survey was conducted are no longer available. Under these conditions, gathering haplotype frequencies involves building large databases and often reshuffling them [Roewer et al., 2020].

2.2.2 Lack of a unique model for calculating haplotype frequencies

Several methods have been described to transform the interrogation of a database into a haplotype-frequency: The *Counting Method*, with or without the application of a ‘*K Correction*’, the *Discrete Laplace Distribution*, and the *simulation method*.

The Counting Method and K Correction

The most elementary method to assess the probability (p) of Y-STR profiles consists of counting the number of times x , in which the observed haplotype occurs in a database of N haplotypes. This approach is called the *Counting Method*.

$$p = \frac{x}{N}$$

The database size employed in this method significantly impacts the accuracy of estimating the rarity of a Y-STR profile. A more significant number of unrelated individuals in the database results in more reliable statistics for calculating a random match frequency estimate.

C.H. Brenner modified this original method by introducing the *Augmented Counting Method* [Brenner, 2010], as follows:

$$p = \frac{x + 1}{N + 1}$$

This author also noticed that when examining Y-STR haplotypes built on numerous loci, an over-abundance of *singletons* arises (singletons are haplotypes observed only once in the database). He recommended adjusting the match probability of singletons by a factor of k , representing the ratio of singletons to non-singletons [Herbert E. Robbins, 1968].

The *k Correction Method* calculates:

$$p = \frac{1 - k}{N + 1}$$

where

$$k = \frac{\alpha + 1}{N + 1}$$

with α representing the total number of singletons in the reference database.

The properties of this statistic were studied by [Buckleton et al., 2011] and [Andersen et al., 2013a].

Discrete Laplace Method

The *Discrete Laplace Method* is a family function in a generalized linear model [Andersen et al., 2013b].

This method starts from the assumption that a population comprises subpopulations, each descending from an ancestral haplotype known as the "*central haplotype*." The central haplotype is supposed to be the most common haplotype within its cluster. The method calculates the probability of a haplotype in a cluster as the product of allele probabilities at each locus. Parameters in the discrete Laplace model describe the spread of repeat lengths of observed alleles, with locus and subpopulation effects. The locus effect reflects mutation rates, and the subpopulation effect reflects subpopulation age, assuming older subpopulations have more chances to mutate from a central haplotype.

The discrete Laplace method has some significant limitations. It assumes Y-STR reference profiles consist of precisely one allele at each locus with an integer repeat

number. This limitation prevents its application to loci like DYS385 a/b and DYF387S1, where most males have exactly two alleles. Additionally, non-integer alleles at low but non-negligible frequencies at several loci in various populations further hinder the method's application [Kruijver et al., 2023].

The underlying model is based on three genetic assumptions [Taylor et al., 2018]:

- i. The collection of haplotypes forms populations that consist of distinct groups known as clades;
- ii. Each clade originates from a common ancestral haplotype through a series of stepwise mutations;
- iii. Mutations take place autonomously and do not interfere with one another.

The Andersen and Balding Method

The complexity and high mutation rate of multi-locus haplotypes in contemporary populations pose challenges in obtaining significant information about the proportion of haplotypes from a sample. Andersen et al. [Andersen & Balding, 2017, 2019] introduce a method that models the distribution of matching males in a population, explicitly emphasizing patrilineal relationships. Their simulation incorporates key factors such as locus mutation rates, population growth rate, and variance in reproductive success. Calibration of the simulation model is performed using observations of 0, 1, or 2 copies of the relevant haplotype.

2.2.3 Assembling pairs, triplets, and quadruplets as haplotypes

Defining all possible combinations that result from a mixed profile pass by constructing pairs, triplets, quadruplets, etc. (in cases of mixtures with two, three, and four contributors, respectively) of haplotypes.

For example, considering a DNA mixture with two contributors and a haplotype of only two loci, four combinations are possible; if are considered four loci, 16 combinations have to be considered; and if the minimum haplotype (9 loci) is considered, the possible haplotype combinations are 512. Going even further, considering the average number of markers included in a commercial Y-STRs amplification kit, i.e. 23, the number of possible haplotype combinations explodes

to 8.388.608. This happens because the haplotype combination increment follows the 2^n rule, where n stays for the number of considered loci.

The formulation of all possible haplotype combinations is further complicated when three or four contributors are considered instead of two, as the number of haplotype combinations will increase exponentially.

2.2.4 Not all conceivable haplotypes are real

Haplotypes envisioned by the combinatorial process are meant to reproduce all haplotypes that can exist according to a free combinatorial process that recalls a random process.

The biological counterpart of this process is the recombination of long-distance genes with virtually no linkage disequilibrium. However, the existence of most of them can legitimately be questioned (Table 1).

Moreover, because there is no automatic way to determine the population frequency of each haplotype in the existing databases, finding out whether or not each combinatorial haplotype has been observed in an actual population becomes a painstaking procedure.

	DYS19	DYS389 I	DYS389 II	DYS390	DYS391	DYS392	DYS393	DYS385	Note
Haplotype 1	16	12	28	21	10	11	14	13,15	Real haplotype
Haplotype 2	9	12	28	21	10	11	14	13,15	Found no match in 349.750 haplotypes

Tab. 1 Example of a haplotype that being different from the real one for just one allele does not exist.

2.2.5 No developed strategies

Solutions for the deconvolution of mixtures of Y-STRs are still practically non-existent or not yet accepted and used by the international scientific community.

2.2.6 Low computing power of personal computers

Constructing quantitative Bayesian LR_s based on Y-mix data requires a computational power that exceeds that of current personal computers. Usually, a 15-loci-based haplotype within a three-person mixture can saturate the computational power of an Intel I₁₀ processor with 554 GB random access memory of the equivalent.

This is a severe limitation to the development of models based on free combinations of Y-STR loci

REFERENCES

- Andersen, M. M., & Balding, D. J. (2017). How convincing is a matching Y-chromosome profile? *PLoS Genetics*, *13*(11). <https://doi.org/10.1371/journal.pgen.1007028>
- Andersen, M. M., & Balding, D. J. (2019). Y-profile evidence: Close paternal relatives and mixtures. *Forensic Science International: Genetics*, *38*, 48–53. <https://doi.org/10.1016/j.fsigen.2018.10.004>
- Andersen, M. M., Eriksen, P. S., & Morling, N. (2013a). *A gentle introduction to the discrete Laplace method for estimating Y-STR haplotype frequencies*. <http://arxiv.org/abs/1304.2129>
- Andersen, M. M., Eriksen, P. S., & Morling, N. (2013b). The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. *Journal of Theoretical Biology*, *329*, 39–51. <https://doi.org/10.1016/j.jtbi.2013.03.009>
- Approved, S. (2010). SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.
- Benschop, C. C. G., van de Merwe, L., de Jong, J., Vanvooren, V., Kempenaers, M., (Kees) van der Beek, C. P., Barni, F., Reyes, E. L., Moulin, L., Pene, L., Haned, H., & Sijen, T. (2017). Validation of SmartRank: A likelihood ratio software for searching national DNA databases with complex DNA profiles. *Forensic Science International: Genetics*, *29*, 145–153. <https://doi.org/10.1016/J.FSIGEN.2017.04.008>
- Bieber, F. R., Buckleton, J. S., Budowle, B., Butler, J. M., & Coble, M. D. (2016). Evaluation of forensic DNA mixture evidence: Protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. *BMC Genetics*, *17*(1). <https://doi.org/10.1186/s12863-016-0429-7>
- Bille, T., Bright, J. A., & Buckleton, J. (2013). Application of Random Match Probability Calculations to Mixed STR Profiles. *Journal of Forensic Sciences*, *58*(2), 474–485. <https://doi.org/10.1111/1556-4029.12067>
- Bille, T. W., Weitz, S. M., Coble, M. D., Buckleton, J., & Bright, J. A. (2014). Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *ELECTROPHORESIS*, *35*(21–22), 3125–3133. <https://doi.org/10.1002/ELPS.201400110>
- Bleka, Ø., Just, R., Agudo, M. M., & Gill, P. (2022). MPSproto: An extension of EuroForMix to evaluate MPS-STR mixtures. *Forensic Science International: Genetics*, *61*, 102781. <https://doi.org/10.1016/J.FSIGEN.2022.102781>
- Brenner, C. H. (2010). Fundamental problem of forensic mathematics - The evidential value of a rare haplotype. *Forensic Science International: Genetics*, *4*(5), 281–291. <https://doi.org/10.1016/j.fsigen.2009.10.013>
- Buckleton, J. S., Krawczak, M., & Weir, B. S. (2011). The interpretation of lineage markers in forensic DNA testing. *Forensic science international. Genetics*, *5*(2), 78–83. <https://doi.org/10.1016/j.fsigen.2011.01.010>

- Butler, J. M., Iyer, H., Press, R., Taylor, M. K., Vallone, P. M., & Willis, S. (2021). *DNA Mixture Interpretation: 5 A NIST Scientific Foundation Review*. <https://doi.org/10.6028/NIST.IR.8351-draft>
- Butler, & John M. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*.
- Costa, C., Figueiredo, C., Amorim, A., Costa, S., Ferreira, P. M., & Pinto, N. (2022). Quantification of forensic genetic evidence: Comparison of results obtained by qualitative and quantitative software for real casework samples. *Forensic Science International: Genetics*, 59. <https://doi.org/10.1016/j.fsigen.2022.102715>
- Cowell, R. G., Graverson, T., Lauritzen, S., & Mortera, J. (2013). *Analysis of Forensic DNA Mixtures with Artefacts*. <http://arxiv.org/abs/1302.4404>
- Cowell, R. G., Lauritzen, S. L., & Mortera, J. (2007). A gamma model for DNA mixture analyses. In *Bayesian Analysis* (Vol. 2, Issue 2). <http://www.staff.city.ac.uk/~rgc/>
- Egeland, T., Dalen, I., & Mostad, P. F. (2003). Estimating the number of contributors to a DNA profile. *International Journal of Legal Medicine*, 117(5), 271–275. <https://doi.org/10.1007/s00414-003-0382-7>
- Evett, I. W., Buffery, C., Willott, G., & Stoney, D. (1991). A guide to interpreting single locus profiles of DNA mixtures in forensic cases. In *Journal of the Forensic Science Society* (Vol. 31, Issue 1, pp. 41–47). [https://doi.org/10.1016/S0015-7368\(91\)73116-2](https://doi.org/10.1016/S0015-7368(91)73116-2)
- Evett, I. W., Gill, P. D., & Lambert, J. A. (1998). *Taking Account of Peak Areas when Interpreting Mixed DNA Profiles*.
- Evett, I. W., & Weir, B. S. (1998). *INTERPRETING DNA EVIDENCE Statistical Genetics for Forensic Scientists*.
- Gill, P. (2019). Interpretation continues to be the main weakness in criminal justice systems: Developing roles of the expert witness and court. *WIREs Forensic Science*, 1(2). <https://doi.org/10.1002/WFS2.1321>
- Gill, P., Benschop, C., Buckleton, J., Bleka, Ø., & Taylor, D. (2021). A review of probabilistic genotyping systems: Euroformix, dnastatistx and strmix™. *Genes*, 12(10), 1559. <https://doi.org/10.3390/GENES12101559/S1>
- Gill, P., Brenner, C. H., Buckleton, J. S., Carracedo, A., Krawczak, M., Mayr, W. R., Morling, N., Prinz, M., Schneider, P. M., & Weir, B. S. (2006). DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International*, 160(2–3), 90–101. <https://doi.org/10.1016/j.forsciint.2006.04.009>
- Gill, P., & Haned, H. (2013). A new methodological framework to interpret complex DNA profiles using likelihood ratios. *Forensic Science International: Genetics*, 7(2), 251–263. <https://doi.org/10.1016/J.FSIGEN.2012.11.002>
- Gill, P., Kirkham, A., & Curran, J. (2007). LoComatioN: A software tool for the analysis of low copy number DNA profiles. *Forensic Science International*, 166(2–3), 128–138. <https://doi.org/10.1016/J.FORSCIINT.2006.04.016>

- Graversen, T., & Lauritzen, S. (2013). Estimation of parameters in DNA mixture analysis. *Journal of Applied Statistics*, 40(11), 2423–2436. <https://doi.org/10.1080/02664763.2013.817549>
- Haned, H. (2010). *Evaluation of statistical methods for the analysis of forensic DNA mixtures*. <https://theses.hal.science/tel-00817181>
- Haned, H. (2011). Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics. *Forensic Science International: Genetics*, 5(4), 265–268. <https://doi.org/10.1016/J.FSIGEN.2010.03.017>
- Haned, H., Benschop, C. C. G., Gill, P. D., & Sijen, T. (2015). Complex DNA mixture analysis in a forensic context: Evaluating the probative value using a likelihood ratio model. *Forensic Science International: Genetics*, 16, 17–25. <https://doi.org/10.1016/J.FSIGEN.2014.11.014>
- Haned, H., Slooten, K., & Gill, P. (2012). Exploratory data analysis for the interpretation of low template DNA mixtures. *Forensic Science International: Genetics*, 6(6), 762–774. <https://doi.org/10.1016/J.FSIGEN.2012.08.008>
- Herbert E. Robbins. (1968). Estimating the Total Probability of the Unobserved Outcomes of an Experiment. *Ann. Math. Statist.* 39 (1) 256-257. <https://doi.org/10.1214/aoms/1177698526>
- John S. Buckleton, Jo-Anne Bright, & Duncan Taylor. (2016). *Forensic DNA Evidence Interpretation* (Second Edition).
- Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., ... Roewer, L. (1997). Evaluation of Y-chromosomal STRs: A multicenter study. *International Journal of Legal Medicine*, 110(3), 125–133. <https://doi.org/10.1007/S004140050051/METRICS>
- Kruijver, M., Taylor, D., & Buckleton, J. (2023). Extending the discrete Laplace method: incorporating multi-copy loci, partial repeats and null alleles. *Forensic Science International: Genetics*, 65. <https://doi.org/10.1016/j.fsigen.2023.102876>
- Lauritzen, S. L., & Mortera, J. (2002). Bounding the number of contributors to mixed DNA stains. *Forensic science international*, 130(2-3), 125–126. [https://doi.org/10.1016/s0379-0738\(02\)00351-1](https://doi.org/10.1016/s0379-0738(02)00351-1)
- Perlin, M. W., & Sinelnikov, A. (2009). An Information Gap in DNA Evidence Interpretation. *PLoS ONE*, 4(12), 8327. <https://doi.org/10.1371/journal.pone.0008327>
- Roewer, L., Andersen, M. M., Ballantyne, J., Butler, J. M., Caliebe, A., Corach, D., D'Amato, M. E., Gusmão, L., Hou, Y., de Knijff, P., Parson, W., Prinz, M., Schneider, P. M., Taylor, D., Vennemann, M., & Willuweit, S. (2020). DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis. In *Forensic Science International: Genetics* (Vol. 48). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.fsigen.2020.102308>

- Schneider, P. M., Fimmers, R., Keil, W., Molsberger, G., Patzelt, D., Pflug, W., Rothämel, T., Schmitter, H., Schneider, H., & Brinkmann, B. (2009). The German Stain Commission: Recommendations for the interpretation of mixed stains. *International Journal of Legal Medicine*, *123*(1), 1–5. <https://doi.org/10.1007/S00414-008-0244-4/METRICS>
- Taylor, D., Bright, J. A., & Buckleton, J. (2013). The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics*, *7*(5), 516–528. <https://doi.org/10.1016/J.FSIGEN.2013.05.011>
- Taylor, D., Curran, J., & Buckleton, J. (2018). Likelihood ratio development for mixed Y-STR profiles. *Forensic Science International: Genetics*, *35*, 82–96. <https://doi.org/10.1016/j.fsigen.2018.03.006>
- Weir, R. S., Triggs, B. S., Starling, C. M., Stowell, L. L., Walsh, L. I., & Buckleton, K. (1997). Interpreting DNA mixtures. *J Forensic Sci*, *42*(2).
- Yang, J., Lin, D., Deng, C., Li, Z., Pu, Y., Yu, Y., Li, K., Li, D., Chen, P., & Chen, F. (2019). The advances in DNA mixture interpretation. *Forensic Science International*, *301*, 101–106. <https://doi.org/10.1016/J.FORSCIINT.2019.05.024>

CHAPTER 3

AIM OF THE PROJECT 48

AIM OF THE PROJECT

In this study, we will illustrate a computational procedure to extract the Value of Evidence (WoE) embedded in biological mixtures whose corresponding Evidence Y (E_Y) is explained by assuming that it originates from only two contributors.

We will, therefore, build a model based on a standard genotype permutation matrix containing 10 Y-markers assigned by current typing procedures, which ends up consisting of two columns reflecting the number of putative contributors and 1024 records mirroring all possible states of the evidence.

The matrix will be designed to be universal and then applicable to any assortment of markers.

After the matrix development, a calculation method will be devised by assuming measurable quantities (typically: peak heights or peak areas) and haplotype frequencies as the baseline data for the computation.

The calculation prototype illustrated in this work will first assume ideal Y-STR haplotypes without allelic drop-outs (modelling drop-outs will be the aim of our future research). As it will be evident, the computation scheme will be sufficiently general to incorporate and model the drop-out itself.

First, we will forge formulas reflecting the ideal ratio at which two contributors intermingle and the rate at which each haplotype state deviates from this ideal ratio. Second, we will calculate the complement to chi-square values ($1-\chi^2$) between the various mix ratios expressed by each theoretically possible haplotype.

Third, we will multiply the peak height ratios, ($1-\chi^2$) values and the haplotype frequency into a single value corresponding to each haplotype's entire probability of existence.

Practical calculation examples will be presented to test the effectiveness of our method.

CHAPTER 4

MATERIALS AND METHODS	50
4.1 Male DNA samples	50
4.2 Experimental male DNA mixtures	50
4.3 Male in silico DNA mock mixtures	50
4.4 Construction of the universal matrix	51
4.5 Calculation worksheet	54
4.6 Statistical Analysis	54
4.6.1 Calculation of the MRs	54
4.6.2 Calculation of the $(1-\chi^2)$ values	55
4.6.3 Calculation of intra-individual PhR	56
4.6.4 Population probabilities	58
4.6.5 Combining the set of evidence in a likelihood	60
4.6.6 Calculation of the Likelihood Ratio	61
4.7 Building worksheets to enable calculations	61
4.8 Data on which to give proof of principle	62
REFERENCES	64

MATERIALS AND METHODS

4.1 Male DNA samples

After informed consent, we asked ten unrelated male individuals to donate nucleated cells in the form of buccal swabs. After collection, samples were immediately anonymized.

Genomic DNA was extracted from each sample using the QIAamp DNA Investigator Kit (Qiagen) and quantified using the 7500 Real-Time PCR System with the Investigator Quantiplex Pro Kit (Applied Biosystems). All samples were then normalized to 1 ng/ μ l.

Samples were individually amplified with the Yfiler™ Plus PCR Amplification Kit (Applied Biosystems); subsequently, amplicons were genotyped using the Capillary Electrophoresis (CE) SeqStudio™ Genetic Analyzer for HID, and data was analyzed using the GeneMapper® ID-X Software v1.6 (Thermo Fisher).

Each procedure was performed strictly following the protocol defined by the manufacturer.

4.2 Experimental male DNA mixtures

DNA samples belonging to two individuals (Ind. 1 and Ind. 2) were thoroughly mixed according to specific proportions. Five different mixture ratios were addressed, starting from a balanced ratio (i.e: the ratio of the amount of DNA from Ind. 1 to that of Ind. 2 is 1:1) and proceeding toward unbalancing according to the following scheme:

- a) 1:1
- b) 1:1.5
- c) 1:3
- d) 1:5
- e) 1:10

4.3 Male in silico DNA mock mixtures

Measurement and dilution errors may affect in silico mixtures much more than commonly believed.

This can compromise our ability to thoroughly test all mixing ratios of our interest.

The following procedure has been adopted as a convenient way to circumvent this problem:

- 1) Two freshly extracted DNA samples belonging to as many unrelated male individuals were accurately quantified by real-time PCR – the quantification was repeated three times, and the three values were averaged. Then, appropriate volumes of both DNA samples were mixed in a test tube in such a way as to reproduce (as far as possible) a one-to-one ratio of total DNA content in the final test tube; the resulting sample was amplified with the Yfiler™ Plus PCR Amplification Kit, run into the SeqStudio™ Genetic Analyzer for HID and the relevant peak heights were noted in an Excel spreadsheet;
- 2) All RFU counts of loci with two peaks were selected, while those with only one peak were discarded;
- 3) The original value corresponding to each locus's minor peak height component was serially altered to reproduce purely fictitious values that exactly produced the mixture ratios of interest;
- 4) The resulting RFU pair value arrays were introduced into the computational matrix and used to predict the WoE.

4.4 Construction of the universal matrix

A 2^{10} matrix of universal adoption allocating whatever RFU pair values of interest was built as follows.

On the upper left-hand side of an Excel worksheet (A1:G11), the called alleles and the RFU values of the 10-loci mock mixture of interest were placed, as shown in Figure 7.

	A	B	C	D	E	F	G
1	Locus	Allele 1	Peak Height 1 (RFU)	Allele 2	Peak Height 2 (RFU)		N. loci
2	DYS576						1
3	DYS389II						2
4	DYS627						3
5	DYS458						4
6	DYS19						5
7	DYS448						6
8	DYS391						7
9	DYS390						8
10	DYS438						9
11	DYS392						10
12							
13		Enter your values in cells B2:E11					
14							

Fig. 7 Empty scheme of the 2¹⁰ matrix of universal adoption allocating whatever RFU pair values of interest.

A 2x2 permutation matrix applying to the initial two pair loci was built according to the following scheme (Figure 8):

- a,b – c,d
- a,d – c,b
- c,b – a,d
- c,d – a,b

and the resulting four haplotype pairs in terms of RFU were:

- 1554,672 – 1493,645
- 1554,645 – 1493,672
- 1493,672 – 1554,645
- 1493,645 – 1554,672

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Locus	Allele 1	Peak Height 1 (RFU)	Allele 2	Peak Height 2 (RFU)		N. loci					Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)
2	DYS576	a	1554	c	1493		1				4	a,b	1554,672	c,d	1493,645
3	DYS389II	b	672	d	645		2				2	a,d	1554,645	c,b	1493,672
4	DYS627						3					c,b	1493,672	a,d	1554,645
5	DYS458						4					c,d	1493,645	a,b	1554,672
6	DYS19						5								
7	DYS448						6								
8	DYS391						7								
9	DYS390						8								
10	DYS438						9								
11	DYS392						10								
12															
13		Enter your values in cells B2:E11													
14															

* yellow: number of permutations
** green: number of loci

Fig. 8 Example of the 2x2 permutation matrix (two contributors, two loci).

To these initial pairs of haplotypes, a third genotype pair was attached in a permutational fashion to obtain the following eight haplotype pairs (Figure 9):

- a,b,e – c,d,f
- a,b,f – c,d,e
- a,d,e – c,b,f
- a,d,f – c,b,e
- c,b,e – a,d,f
- c,b,f – a,d,e
- c,d,e – a,b,f
- c,d,f – a,b,e

	A	B	C	D	E	F	G	H	I	J	P	Q	R	S	T	
1	Locus	Allele 1	Peak Height 1 (RFU)	Allele 2	Peak Height 2 (RFU)	N. loci						Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	
2	DYS576	a	1554	c	1493	1						a,b,e	1554,672,1032	c,d,f	1493,645,991	
3	DYS389II	b	672	d	645	2						a,b,f	1554,672,991	c,d,e	1493,645,1032	
4	DYS627	e	1032	f	991	3						a,d,e	1554,645,1032	c,b,f	1493,672,991	
5	DYS458					4						a,d,f	1554,645,991	c,b,e	1493,672,1032	
6	DYS19					5						c,b,e	1493,672,1032	a,d,f	1554,645,991	
7	DYS448					6						c,b,f	1493,672,991	a,d,e	1554,645,1032	
8	DYS391					7						c,d,e	1493,645,1032	a,b,f	1554,672,991	
9	DYS390					8						c,d,f	1493,645,991	a,b,e	1554,672,1032	
10	DYS438					9										
11	DYS392					10										
12																
13		Enter your values in cells B2:E11														

Fig. 9 Example of the 2x3 permutation matrix (two contributors, three loci).

This process of permutational attachment was repeated up to the tenth locus RFU pair, and the resulting matrix corresponds to Figure 10:

	A	B	C	D	E	F	G	H	AV	AZ	BA	BB	BC	BD		
1	Locus	Allele 1	Peak Height 1 (RFU)	Allele 2	Peak Height 2 (RFU)	N. loci						Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	
2	DYS576	a	1554	c	1493	1						a,b,e.g.i.k.m.o.q.s	1554,672,1032,1558,1475,1028,1190,1064,1172	c,d,f,h,j.l.n.p.r.t	1493,645,991,1497,1417,987,1143,1022,1126,94	
3	DYS389II	b	672	d	645	2						a,b,e.g.i.k.m.o.q.t	1554,672,1032,1558,1475,1028,1190,1064,1172	c,d,f,h,j.l.n.p.r.s	1493,645,991,1497,1417,987,1143,1022,1126,98	
4	DYS627	e	1032	f	991	3						a,b,e.g.i.k.m.o.r.s	1554,672,1032,1558,1475,1028,1190,1064,1126	c,d,f,h,j.l.n.p.q.t	1493,645,991,1497,1417,987,1143,1022,1172,94	
5	DYS458	g	1558	h	1497	4						a,b,e.g.i.k.m.o.r.t	1554,672,1032,1558,1475,1028,1190,1064,1126	c,d,f,h,j.l.n.p.q.s	1493,645,991,1497,1417,987,1143,1022,1172,98	
6	DYS19	i	1475	j	1417	5						a,b,e.g.i.k.m.p.q.s	1554,672,1032,1558,1475,1028,1190,1022,1172	c,d,f,h,j.l.n.o.r.t	1493,645,991,1497,1417,987,1143,1064,1126,94	
7	DYS448	k	1028	l	987	6						a,b,e.g.i.k.m.p.q.t	1554,672,1032,1558,1475,1028,1190,1022,1172	c,d,f,h,j.l.n.o.r.s	1493,645,991,1497,1417,987,1143,1064,1126,98	
8	DYS391	m	1190	n	1143	7						a,b,e.g.i.k.m.p.r.s	1554,672,1032,1558,1475,1028,1190,1022,1126	c,d,f,h,j.l.n.o.q.t	1493,645,991,1497,1417,987,1143,1064,1172,94	
9	DYS390	o	1064	p	1022	8						a,b,e.g.i.k.m.p.r.t	1554,672,1032,1558,1475,1028,1190,1022,1126	c,d,f,h,j.l.n.o.q.s	1493,645,991,1497,1417,987,1143,1064,1172,98	
10	DYS438	q	1172	r	1126	9						a,b,e.g.i.k.n.o.q.s	1554,672,1032,1558,1475,1028,1143,1064,1172	c,d,f,h,j.l.m.p.r.t	1493,645,991,1497,1417,987,1190,1022,1126,94	
11	DYS392	s	985	t	946	10						a,b,e.g.i.k.n.o.q.t	1554,672,1032,1558,1475,1028,1143,1064,1172	c,d,f,h,j.l.m.p.q.s	1493,645,991,1497,1417,987,1190,1022,1172,94	
12												a,b,e.g.i.k.n.o.r.s	1554,672,1032,1558,1475,1028,1143,1064,1126	c,d,f,h,j.l.m.p.q.t	1493,645,991,1497,1417,987,1190,1022,1172,94	
13		Enter your values in cells B2:E11														
14												a,b,e.g.i.k.n.p.q.s	1554,672,1032,1558,1475,1028,1143,1022,1172	c,d,f,h,j.l.m.o.r.t	1493,645,991,1497,1417,987,1190,1064,1126,94	
15																
16												c,d,f,h,j.l.n.o.q.t	1493,645,991,1497,1417,987,1143,1064,1172,94	a,b,e.g.i.k.m.p.r.s	1554,672,1032,1558,1475,1028,1190,1022,1126	
17												c,d,f,h,j.l.n.o.r.s	1493,645,991,1497,1417,987,1143,1064,1126,98	a,b,e.g.i.k.m.p.q.t	1554,672,1032,1558,1475,1028,1190,1022,1172	
18												c,d,f,h,j.l.n.o.r.t	1493,645,991,1497,1417,987,1143,1064,1126,94	a,b,e.g.i.k.m.p.q.s	1554,672,1032,1558,1475,1028,1190,1022,1172	
19												c,d,f,h,j.l.n.p.q.s	1493,645,991,1497,1417,987,1143,1022,1172,98	a,b,e.g.i.k.m.o.r.t	1554,672,1032,1558,1475,1028,1190,1064,1126	
20												c,d,f,h,j.l.n.p.q.t	1493,645,991,1497,1417,987,1143,1022,1172,94	a,b,e.g.i.k.m.o.r.s	1554,672,1032,1558,1475,1028,1190,1064,1126	
21												c,d,f,h,j.l.n.p.r.s	1493,645,991,1497,1417,987,1143,1022,1126,98	a,b,e.g.i.k.m.o.q.t	1554,672,1032,1558,1475,1028,1190,1064,1172	
22												c,d,f,h,j.l.n.p.r.t	1493,645,991,1497,1417,987,1143,1022,1126,94	a,b,e.g.i.k.m.o.q.s	1554,672,1032,1558,1475,1028,1190,1064,1172	

Fig. 10 Example of the 2x10 permutation matrix (two contributors, ten loci).

This matrix is of universal usage insofar as whatever allele name/RFU of interest can be pasted in A1-G11, and the final permutation is automatically updated.

Intermediate matrices from 2 to 9 genotype pairs can be used as autonomous matrices whenever it is necessary to intervene on each of them.

4.5 Calculation worksheet

The 1024 record’s matrix is pasted into a fresh Excel spreadsheet, and the permuted values are parsed into strings of individual values while preserving the permutational scheme (Figure 11).

The image shows a partial representation of a 2¹⁰ matrix. The spreadsheet has columns labeled A through X. Column A contains a list of 1024 records, each with a unique identifier and a genotype string (e.g., 'a.b.c.d.e.f.g.h.i.j.k.l.m.n.o.p.q.r.s.t.u.v.w.x'). Columns B through X contain numerical data for each record, representing peak heights and other calculated values. The data is organized into several groups, with some columns highlighted in green and others in yellow. The values are generally small integers, ranging from 0 to 1000.

Fig. 11 Partial representation of a 2¹⁰ matrix with data divided into columns.

4.6 Statistical Analysis

4.6.1 Calculation of the MRs

We now have 1024 records subdivided into two series of assorted values as afforded by the permutational scheme of the 2-person matrix (PM) model.

At each record, the average RFU values of individuals 1 and 2 were calculated, and the quantitative ratio across the two individuals was drawn.

$$MR1 = \frac{\text{Peak height of Ind. 1 at locus 1}}{\text{Sum of Peak heights of Ind. 1 and 2 at locus 1}}$$

To give an example, at record 1 of the previous permutation, the two individuals’ haplotypes gave rise to the following pair of average RFU and MR1 values (Table 2).

Ind. 1 Peak Height (RFU)										Ind. 2 Peak Height (RFU)										Avg RFU Ind.1	Avg RFU Ind.2	MR1	MR1
1554	672	1032	1558	1475	1028	1190	1064	1172	985	284	68	92	252	141	111	108	87	107	73	1173	132	0,845484	0,154516

Tab. 2 Average RFU and MR1 values of the first record of the 2PM mixture taken as an example.

This refers to locus 1. The same calculation must be reproduced for each locus, determining MR2, MR3, etc. (Table 3).

Avg RFU Ind.1	Avg RFU Ind.2	MR1	MR1	MR2	MR2	MR3	MR3	MR4	MR4	MR5	MR5	MR6	MR6	MR7	MR7	MR8	MR8	MR9	MR9	MR10	MR10
1173	132	0,8455	0,1545	0,9081	0,0919	0,9181	0,0819	0,8608	0,1392	0,9127	0,0873	0,9025	0,0975	0,9168	0,0832	0,9244	0,0756	0,9163	0,0837	0,9310	0,0690

Tab. 3 Average RFU and all the MR values of the first record of the 2PM mixture taken as an example.

4.6.2 Calculation of the $(1-\chi^2)$ values

At each record, ten different MR values can be calculated from the individual RFU values originating to every single locus. Once calculated, each locus-specific MR is compared to the average MR value by using the $(1-\chi^2)$ (Figure 12):

$$(1 - \chi^2) = 1 - \frac{(obs - exp)^2}{exp}$$

	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD
1	MR1	MR1	MR2	MR2	MR3	MR3	MR4	MR4	MR5	MR5	MR6	MR6	MR7	MR7	MR8	MR8	MR9	MR9	MR10	MR10	Avg MR Ind.1	Avg MR Ind.2	x ²	x ²	x ²	x ²	x ²	x ²	x ²	x ²	x ²	x ²
2	0,84548	0,15452	0,90811	0,09189	0,91815	0,08185	0,86077	0,13923	0,91275	0,08725	0,90255	0,09745	0,9168	0,0832	0,92441	0,07559	0,91634	0,08366	0,931	0,069	0,90364	0,09636	0,84378	0,98791	0,96077	0,88451	0,97537	0,99705	0,86443	0,84387	0,85666	0,92611

Fig. 12 Calculation of the $(1-\chi^2)$ values for the first record of the 2PM mixture taken as an example.

To better comprehend Figure 12, we reproduce the same results in Table 4, which illustrates the results of the $(1-\chi^2)$ calculation for the first record.

MR Ind. 1	MR Ind. 2	Avg MR Ind.1	Avg MR Ind.2	χ^2
0,8455	0,1545	0,9036	0,0964	0,8438
0,9081	0,0919	0,9036	0,0964	0,9879
0,9181	0,0819	0,9036	0,0964	0,9608
0,8608	0,1392	0,9036	0,0964	0,8845
0,9127	0,0873	0,9036	0,0964	0,9754
0,9025	0,0975	0,9036	0,0964	0,9971
0,9168	0,0832	0,9036	0,0964	0,9644
0,9244	0,0756	0,9036	0,0964	0,9439
0,9163	0,0837	0,9036	0,0964	0,9657
0,9310	0,0690	0,9036	0,0964	0,9261

Tab. 4 Calculation of the $(1-\chi^2)$ values for the first record of the 2PM mixture taken as an example.

When frequencies are available, the Chi-square is usually used to test hypotheses about the distribution of observations in different categories.

We had two series of data: the experimental frequency data and the expected frequency data.

The null hypothesis to test with χ^2 is that the observed frequencies are the same as the expected frequencies (except for chance variation). If the observed and expected frequencies are the same, then the χ^2 value is zero.

If the frequencies we observe differ from those expected, the value of χ^2 increases.

The larger the value of χ^2 , the more significant the difference between observed and expected (In the formula, observed minus expected values are squared to avoid the negative sign).

Consequently, while using $(1 - \chi^2)$, a higher value will reflect a close resemblance between the observed and expected data series.

In the Excel worksheet, several χ^2 formulas are available. If using “TEST.CHI.QUAD” the relevant value obtained is equivalent to $(1 - \chi^2)$. We used the “TEST.CHI.QUAD” formula as a likelihood value.

At the end of this computational step, the chi-square values for each of the 1024 records (or states) are obtained.

Finally, the ten values are multiplied together across each record.

4.6.3 Calculation of intra-individual PhR

At autosomal loci, the PhR [Kelly et al., 2012] measures the genotype plausibility of every heterozygous genotype. In its simplest form, this index is calculated by dividing the minor height allele by the major.

Y genotypes have no PhR because they have just one allele per locus.

However, based on the following assumptions, throughout this paper, we will lend the acronym ‘PhR’ to indicate a non-conventional index for Y loci elicited at a mixture.

- Let us have – as we indeed have in this experimental design – a 2PM mixture and twenty different RFU values distributed pairwise at 10 loci. Because of the mixed evidence uncertainty, each RFU per each locus may, in principle – indifferently – be assigned to individuals 1 or 2 of the mixture. A process of random assignment of RFU values to either individual would give rise, as we have seen, to 1024 pairs of different assortments. Nevertheless, not all those

assortments are as equally probable. That depends on the quantitative ratio operating among contributors.

- Highly balanced mixtures – or mixtures whose contributors give the same amount of DNA to the relevant race – give rise to nearly equally plausible assortments, as shown in Figure 13:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Locus	Allele 1	Peak Height 1 (RFU)	Allele 2	Peak Height 2 (RFU)	N. loci						Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)
2	DYS576	a	1550	c	1440	1					4	a,b	1550,1225	c,d	1440,1199
3	DYS389II	b	1225	d	1199	2					2	a,d	1550,1199	c,b	1440,1225
4	DYS627					3						c,b	1440,1225	a,d	1550,1199
5	DYS458					4						c,d	1440,1199	a,b	1550,1225
6	DYS19					5									
7	DYS448					6									
8	DYS391					7									
9	DYS390					8									
10	DYS438					9									
11	DYS392					10									
12															
13		Enter your values in cells B2:E11													

* yellow: number of permutations
** green: number of loci

Fig. 13 Example of a highly balanced 2x2 mixture and all the possible assortments.

- Conversely, permutation quantities at unbalanced mixtures would give rise to highly probable and highly improbable permutations, as shown in Figure 14:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Locus	Allele 1	Peak Height 1 (RFU)	Allele 2	Peak Height 2 (RFU)	N. loci						Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)					
2	DYS576	a	1550	c	125	1					4	a,b	1550,1225	c,d	125,102					
3	DYS389II	b	1225	d	102	2					2	a,d	1550,102	c,b	125,1225					→ Highly probable + highly probable haplotypes
4	DYS627					3						c,b	125,1225	a,d	1550,102					→ improbable + improbable haplotypes
5	DYS458					4						c,d	125,102	a,b	1550,1225					→ Highly probable + highly probable haplotypes
6	DYS19					5														
7	DYS448					6														
8	DYS391					7														
9	DYS390					8														
10	DYS438					9														
11	DYS392					10														
12																				
13		Enter your values in cells B2:E11																		

* yellow: number of permutations
** green: number of loci

Fig. 14 Example of an unbalanced 2x2 mixture and all the possible assortments.

A way to determine what state is probable and what is not would be to convert the RFU values to a quantitative ratio (MR) and compare ratio values across the same individual. For the permutation to be true, these values must be close to one another and ideally equal if compensating for error measurements.

Therefore, introducing an intra-individual peak height ratio evaluation should help choose all RFU values belonging to Individual 1 and all those belonging to Individual 2. The number of these particular PhRs that need to be calculated at each record for each contributor at a 2PM 10 loci is 45 (9+8+7+6+5+4+3+2+1= 45) (Figure 15).

Individual 1									
L1 vs L2-L10 (RFU)	L2 vs L3-L10 (RFU)	L3 vs L4-L10 (RFU)	L4 vs L5-L10 (RFU)	L5 vs L6-L10 (RFU)	L6 vs L7-L10 (RFU)	L7 vs L8-L10 (RFU)	L8 vs L9-L10 (RFU)	L9 vs L10 (RFU)	
1554 672	672 1032	1032 1558	1558 1475	1475 1028	1028 1190	1190 1064	1064 1172	1172 985	
1554 1032	672 1558	1032 1475	1558 1028	1475 1190	1028 1064	1190 1172	1064 985		
1554 1558	672 1475	1032 1028	1558 1190	1475 1064	1028 1172	1190 985			
1554 1475	672 1028	1032 1190	1558 1064	1475 1172	1028 985				
1554 1028	672 1190	1032 1064	1558 1172	1475 985					
1554 1190	672 1064	1032 1172	1558 985						
1554 1064	672 1172	1032 985							
1554 1172	672 985								
1554 985									

Individual 1									
L1 vs L2-L10 (PhR)	L2 vs L3-L10 (PhR)	L3 vs L4-L10 (PhR)	L4 vs L5-L10 (PhR)	L5 vs L6-L10 (PhR)	L6 vs L7-L10 (PhR)	L7 vs L8-L10 (PhR)	L8 vs L9-L10 (PhR)	L9 vs L10 (PhR)	
0,4324	0,6512	0,6624	0,9467	0,6969	0,8639	0,8941	0,9078	0,8404	
0,6641	0,4313	0,6997	0,6598	0,8068	0,9662	0,9849	0,9258		
0,9974	0,4556	0,9961	0,7638	0,7214	0,8771	0,8277			
0,9492	0,6537	0,8672	0,6829	0,7946	0,9582				
0,6615	0,5647	0,9699	0,7522	0,6678					
0,7658	0,6316	0,8805	0,6322						
0,6847	0,5734	0,9545							
0,7542	0,6822								
0,6338									

Individual 2									
L1 vs L2-L10 (RFU)	L2 vs L3-L10 (RFU)	L3 vs L4-L10 (RFU)	L4 vs L5-L10 (RFU)	L5 vs L6-L10 (RFU)	L6 vs L7-L10 (RFU)	L7 vs L8-L10 (RFU)	L8 vs L9-L10 (RFU)	L9 vs L10 (RFU)	
284 68	68 92	92 252	252 141	141 111	111 108	108 87	87 107	107 73	
284 92	68 252	92 141	252 111	141 108	111 87	108 107	87 73		
284 252	68 141	92 111	252 108	141 87	111 107	108 73			
284 141	68 111	92 108	252 87	141 107	111 73				
284 111	68 108	92 87	252 107	141 73					
284 108	68 87	92 107	252 73						
284 87	68 107	92 73							
284 107	68 73								
284 73									

Individual 2									
L1 vs L2-L10 (PhR)	L2 vs L3-L10 (PhR)	L3 vs L4-L10 (PhR)	L4 vs L5-L10 (PhR)	L5 vs L6-L10 (PhR)	L6 vs L7-L10 (PhR)	L7 vs L8-L10 (PhR)	L8 vs L9-L10 (PhR)	L9 vs L10 (PhR)	
0,2394	0,7391	0,3651	0,5595	0,7872	0,9730	0,8056	0,8131	0,6822	
0,3239	0,2698	0,6525	0,4405	0,7660	0,7838	0,9907	0,8391		
0,8873	0,4823	0,8288	0,4286	0,6170	0,9640	0,6759			
0,4965	0,6126	0,8519	0,3452	0,7589	0,6577				
0,3908	0,6296	0,9457	0,4246	0,5177					
0,3803	0,7816	0,8598	0,2897						
0,3063	0,6355	0,7935							
0,3768	0,9315								
0,2570									

Fig. 15 Forty-five Intra-individual peak height ratio calculated for the first record for each contributor of the 2PM mixture taken as an example.

4.6.4 Population probabilities

As we have seen, developing a permutation matrix based on ten individual Y loci at 2PM gives rise to 1024 records. Each record is based on a pair of complementary

haplotypes whose combined alleles must account for the whole mixed evidence (E_{mix}).

Overall, to compute a basic likelihood ratio and account for the presence of a given POI in a 2PM mixture, we need to ascertain the frequency of 2048 different haplotypes ($1024 \times 2 = 2048$) from the relevant population data.

There is no other way to determine a haplotype frequency except that of interrogating the YHRD database [Willuweit & Roewer, 2015], whence frequencies need to be extracted individually – one frequency for just every haplotype at the time. This is exceedingly demanding and time-consuming and cannot be quickly done.

Another perplexing aspect of this task is that a very scanty proportion of the 2048 haplotypes that can theoretically be envisioned as present in the permutation may exist and be recovered in the YHRD database. For example, of the 2048 haplotypes built within the matrix in Figure 11 (page 54), only a handful are represented in the most extensive population available to us.

Therefore, this leaves the problem of finding all the population data needed to perform the LR calculation.

A possible way to solve this problem is to determine only the POI haplotype frequency and assume all other haplotypes have the same probability to exist as the POI. This is equivalent to saying that we behave in such a way that the resulting denominator in the LR is overestimated.

In short, frequencies are obtained by interrogating a finite database; nearly all queries are based on theoretical haplotypes, that may not exist at all; then it is unrealistic to expect the sum of the frequencies of the 2048 haplotypes to be equal to 1. The whole assumption turns out to be conservative, as it increases the numerical value placed at the denominator.

While comparing population and quantitative probabilities, one realizes that the impact of whatever database frequencies in themselves becomes negligible.

In the following chapters, in fact, it will become apparent that one such approach does not noticeably modify the final LR value whenever the mixture is quantitatively imbalanced. In all these cases, the populational probability is

negligible when calculating LRs. This somewhat surprising finding was anticipated by Andersen and Balding [Andersen & Balding, 2017]

It is tantamount to saying that in unbalanced mixtures, the LR can be calculated based only on the POI haploid frequency.

4.6.5 Combining the set of evidence in a likelihood

To sum up, we have stuck to three different likelihoods involved in calculating the final LR: haplotype frequencies drawn from a population database, the $(1-\chi^2)$ index and the PhR.

It can be easily demonstrated that the three indexes are independent of each other (Figure 16):

- The haplotype frequencies pertain to a specific population;
- The $(1-\chi^2)$ mirrors an intrinsic property of the particular mixture in the matter and, exactly, the ratio at which the two contributors have mixed in the act of pouring in the same stain;
- The PhR reflects the congruity that has to exist within all DNA quantities distributed to each locus. So, if, for example, a certain individual has contributed one-tenth to the whole amount of DNA contained in a mixed stain, all loci will be represented by alleles that express one-tenth of the RFU signal assigned to the whole locus in the EPG. Therefore, the theoretical PhR within each individual should be 1, except for the error measurements. What is introduced in the quantitative scheme of calculation is the error measurement.

Because the three likelihoods are independent, they can be multiplied.

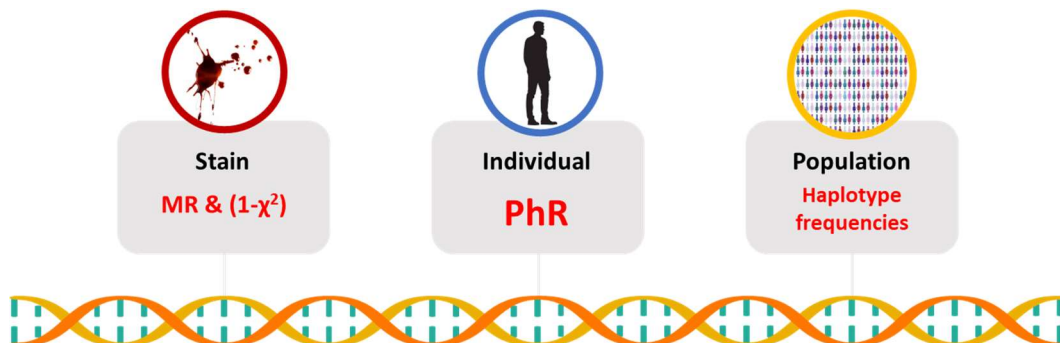


Fig. 16 The three indexes are independent of each other.

4.6.6 Calculation of the Likelihood Ratio

Within a two-person mixture there can be just one meaningful likelihood ratio:

$$LR = \frac{\Pr(E_{mix}|K * U)}{\Pr(E_{mix}|U * U)}$$

where K is the POI and U is an unknown individual.

The **numerator** will be represented by just one likelihood or the only record in the matrix that exactly contains the POI. This record will contain ten $(1-\chi^2)$ values, 90 PhRs and the POI haplotype frequency.

The **denominator** will be the sum of all 1024 likelihoods entering the 2PM matrix, and each record is determined by the product of ten $(1-\chi^2)$, 90 PhRs and the square of the haplotype frequency (Figure 17).

EQ	ER	ES	ET	EU	EV
Haplotype frequency POI	Set of evidence	Numerator	Denominator	LR	
1,30E-05	4,38E-27	1,13E-43	-		
Haplotype Frequency estimated in Yhrd					
Product of $(1-\chi^2)$, PhRs and the square of the haplotype frequency.					
Product of $(1-\chi^2)$, PhRs and the POI haplotype frequency					
Sum of all 1024 likelihoods					

Fig. 17 Explanation of the indexes needed for the calculation of the LR.

This will amount to nearly 500.000 small calculations. Such a large amount of small and, in principle, simple calculations saturate the computational capabilities of the average personal computers available today.

According to our experience, an NVIDIA Intel Core i9 processor can handle calculation worksheets allocating 7 to 8 loci without falling into stall. Ten loci matrices are very difficult to harness by using the same processor.

4.7 Building worksheets to enable calculations

For practical reasons, the matrix is constructed using a dedicated worksheet (see Figure 11 for the 2PM 10 loci matrix).

The matrix worksheet is of universal use, and it is sufficient to paste fresh data in A1:G11 cells to give back the 1024 pairs of permutation.

The updated 1024 matrix is copied and pasted into a calculation spreadsheet (Table 5). There can be as many calculation worksheets for as many loci to include for computing.

Excel sheet cells	Argument
A30:A130	Haplotype of Individual 1
B30:A130	Peak Height of Individual 1' haplotype (RFU)
C30:C130	Haplotype of Individual 2
D30:D130	Peak Height of Individual 2' haplotype (RFU)
E3:K130	Peak heights of Ind.1 placed in column
L3:R130	Peak heights of Ind.2 placed in column
S3:S130	Average of Peak heights of Ind.1
T3:T130	Average of Peak heights of Ind.2
U3:U130	MR of the average of Peak heights of Ind.1
V3:V130	MR of the average of Peak heights of Ind.2
W3:X130	MR locus 1
Y3:Z130	MR locus 2
AA3:AB130	MR locus 3
AC3:AD130	MR locus 4
AE3:AF130	MR locus 5
AG3:AH130	MR locus 6
AI3:AJ130	MR locus 7
AK3:AQ130	χ^2 of loci 1-7
AR3:BL130	Intra-individual peak height ratio of Ind.1
BM3:CG130	Intra-individual peak height ratio of Ind.2
CH1	Haplotype frequency of the POI
CH3:CH130	Product of $(1-\chi^2)$, PhRs and the square of the haplotype frequency
CH2	Denominator = sum of CH3:CH130
CI3:CI130	Numerator = Product of $(1-\chi^2)$, PhRs and the haplotype frequency of the POI
CJ3:CJ130	Likelihood Ratios

Tab. 5 Explanation of each block of the Excel spreadsheet used to calculate the LR.

4.8 Data on which to give proof of principle

As mentioned, working with in silico mixtures does not guarantee that the ratio of interest is addressed because these mixtures are affected by experimental errors in generating the data.

Therefore, we preferred to adopt mock mixtures, allowing the operator to harness the mixing ratio of interest.

The mock mixtures used for this work reflected the following ratios (Table 6):

Mock Mixture		Amount (%)
1	Ind. 1	50,3
	Ind. 2	49,7
2	Ind. 1	51
	Ind. 2	49
3	Ind. 1	52
	Ind. 2	48
4	Ind. 1	53
	Ind. 2	47
5	Ind. 1	54
	Ind. 2	46
6	Ind. 1	55
	Ind. 2	45
7	Ind. 1	56
	Ind. 2	44
8	Ind. 1	57
	Ind. 2	43
9	Ind. 1	58
	Ind. 2	42
10	Ind. 1	59
	Ind. 2	41
11	Ind. 1	60
	Ind. 2	40
12	Ind. 1	61
	Ind. 2	39
13	Ind. 1	62
	Ind. 2	38
14	Ind. 1	63
	Ind. 2	37
15	Ind. 1	64
	Ind. 2	36
16	Ind. 1	65
	Ind. 2	35

Mock Mixture		Amount (%)
17	Ind. 1	66
	Ind. 2	34
18	Ind. 1	67
	Ind. 2	33
19	Ind. 1	68
	Ind. 2	32
20	Ind. 1	69
	Ind. 2	31
21	Ind. 1	70
	Ind. 2	30
22	Ind. 1	71
	Ind. 2	29
23	Ind. 1	72
	Ind. 2	28
24	Ind. 1	73
	Ind. 2	27
25	Ind. 1	74
	Ind. 2	26
26	Ind. 1	75
	Ind. 2	25
27	Ind. 1	76
	Ind. 2	24
28	Ind. 1	77
	Ind. 2	23
29	Ind. 1	78
	Ind. 2	22
30	Ind. 1	79
	Ind. 2	21
31	Ind. 1	80
	Ind. 2	20
32	Ind. 1	91
	Ind. 2	9

Tab. 6 Mock mixtures used in this work and the contribution amount of individuals 1 and 2.

REFERENCES

- Andersen, M. M., & Balding, D. J. (2017). How convincing is a matching Y-chromosome profile? *PLoS Genetics*, *13*(11). <https://doi.org/10.1371/journal.pgen.1007028>
- Kelly, H., Bright, J. A., Curran, J. M., & Buckleton, J. (2012). Modelling heterozygote balance in forensic DNA profiles. *Forensic Science International. Genetics*, *6*(6), 729–734. <https://doi.org/10.1016/J.FSIGEN.2012.08.002>
- Willuweit, S., & Roewer, L. (2015). The new Y Chromosome Haplotype Reference Database. *Forensic Science International. Genetics*, *15*, 43–48. <https://doi.org/10.1016/J.FSIGEN.2014.11.024>

CHAPTER 5

RESULTS	66
5.1 Male DNA samples	66
5.2 Experimental male DNA mixtures	66
5.3 Male in silico DNA mock mixtures	67
5.4 Construction of the universal matrix	69
5.5 Statistical analysis	69
DISCUSSION	72
REFERENCES	79

RESULTS

5.1 Male DNA samples

Only two of the ten male samples were selected according to a specific criterion. In addition to not being related, the two individuals had to exhibit two sufficiently different haplotypes to guarantee a high number of loci in which the allele of Individual 1 did not overlap with that of Individual 2.

Samples 22-028-A and 22-028-B were found to have eight overlapping Y-STRs out of 25 loci. Therefore, the presence of 17 distinct loci provides a reasonable basis for creating a DNA mixture with these two male contributors.

The two haplotypes are represented in Table 7.

	DYS76	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533
22-028-A	19	13	23	30	23	11	18	14	12	19	12	15	25	12	13	37	18	15	14	29	14	12	23	36	13
22-028-B	20	13	23	29	21	11	15	17	12	20	10	15	23	10	11	35	19	15	12	29	12	13	22	37	13

Tab. 7 Haplotypes of Individuals 22-028-A and 22-028-B.

For continuity of reading, Individuals 22-028-A and 22-028-B will be referred to as Individuals/Contributors 1 and 2, respectively.

5.2 Experimental male DNA mixtures

Five mixtures were set up, by changing the DNA concentration of the two individuals within the mixture (1:1, 1:1,5, 1:3, 1:5, 1:10).

Given the genomic DNA extracted from Individuals 1 and 2, which was found to be 0.017 ng/ μ l and 0.14 ng/ μ l, respectively, the DNA mixtures were generated as reported in Table 8:

	DNA Quantity (ng)		Mixture composition (μ l)	
	22-028-A [0,017 ng/ μ l]	22-028-B [0,14 ng/ μ l]	22-028-A [0,017 ng/ μ l]	22-028-B [0,14 ng/ μ l]
1:1	0,051	0,051	9	1,1
1:1,5	0,051	0,077	6	1,1
1:3	0,051	0,153	3	1,1
1:5	0,051	0,255	3	1,8
1:10	0,051	0,510	3	3,6

Tab. 8 DNA composition of the in silico male mixtures.

The attempt to set up and analyze real-world mixtures was made in the early stages of the work before developing our mock mixtures. We were soon discouraged from using real mixtures due to the high error rate of the relevant laboratory procedure. Therefore, we realized that in-silico mock mixtures are by far more convenient to generate and use.

5.3 Male in silico DNA mock mixtures

For the creation of the mock mixtures, only the 17 loci where Individuals 1 and 2 exhibited distinct alleles (DYS576, DYS389II, DYS627, DYS458, DYS19, DYS448, DYS391, DYS390, DYS438, DYS392, DYS570, DYS385, DYS393, DYS439, DYS481, and DYF387S1) were initially considered. However, due to the limited computational capacity of the personal computer used, only seven loci (DYS576, DYS389II, DYS627, DYS458, DYS19, DYS448, DYS391) were effectively investigated for LR calculations in this study.

The resulting RFU pair values arrays are shown in Table 9.

		DYS576	DYS389II	DYS627	DYS458	DYS19	DYS448	DYS391	
		Haplotypes							
Ind.1 22-028-A		19	30	23	18	14	19	12	
Ind.2 22-028-B		20	29	21	15	17	20	10	
Mock Mixture		Peak Height (RFU)							Amount (%)
1	Ind. 1	1554	672	1032	1558	1475	1028	1190	50.3
	Ind. 2	1538	665	1021	1542	1460	1017	1178	49.7
2	Ind. 1	1554	672	1032	1558	1475	1028	1190	51
	Ind. 2	1493	645	991	1497	1417	987	1143	49
3	Ind. 1	1554	672	1032	1558	1475	1028	1190	52
	Ind. 2	1434	620	952	1438	1362	949	1098	48
4	Ind. 1	1554	672	1032	1558	1475	1028	1190	53
	Ind. 2	1398	604	928	1402	1327	925	1071	47
5	Ind. 1	1554	672	1032	1558	1475	1028	1190	54
	Ind. 2	1320	571	877	1324	1253	873	1011	46
6	Ind. 1	1554	672	1032	1558	1475	1028	1190	55
	Ind. 2	1271	550	844	1275	1207	841	974	45
7	Ind. 1	1554	672	1032	1558	1475	1028	1190	56
	Ind. 2	1221	528	811	1224	1159	808	935	44
8	Ind. 1	1554	672	1032	1558	1475	1028	1190	57
	Ind. 2	1165	504	774	1168	1106	771	892	43
9	Ind. 1	1554	672	1032	1558	1475	1028	1190	58
	Ind. 2	1125	486	747	1127	1067	744	861	42
10	Ind. 1	1554	672	1032	1558	1475	1028	1190	59
	Ind. 2	1079	466	716	1082	1024	713	826	41
11	Ind. 1	1554	672	1032	1558	1475	1028	1190	60
	Ind. 2	1035	447	687	1038	983	685	793	40
12	Ind. 1	1554	672	1032	1558	1475	1028	1190	61
	Ind. 2	993	429	659	995	942	656	760	39
13	Ind. 1	1554	672	1032	1558	1475	1028	1190	62
	Ind. 2	1041	409	628	948	898	626	724	38
14	Ind. 1	1554	672	1032	1558	1475	1028	1190	63
	Ind. 2	916	396	608	919	870	606	702	37
15	Ind. 1	1554	672	1032	1558	1475	1028	1190	64
	Ind. 2	878	379	583	880	833	580	672	36
16	Ind. 1	1554	672	1032	1558	1475	1028	1190	65
	Ind. 2	834	360	554	836	792	552	639	35
17	Ind. 1	1554	672	1032	1558	1475	1028	1190	66
	Ind. 2	798	345	530	800	758	528	611	34
18	Ind. 1	1554	672	1032	1558	1475	1028	1190	67
	Ind. 2	761	329	505	763	722	503	583	33
19	Ind. 1	1554	672	1032	1558	1475	1028	1190	68
	Ind. 2	731	316	485	733	694	483	560	32
20	Ind. 1	1554	672	1032	1558	1475	1028	1190	69
	Ind. 2	696	301	462	697	660	460	533	31
21	Ind. 1	1554	672	1032	1558	1475	1028	1190	70
	Ind. 2	665	287	441	666	631	439	509	30
22	Ind. 1	1554	672	1032	1558	1475	1028	1190	71
	Ind. 2	637	275	423	638	604	421	487	29
23	Ind. 1	1554	672	1032	1558	1475	1028	1190	72
	Ind. 2	606	262	402	607	575	400	464	28
24	Ind. 1	1554	672	1032	1558	1475	1028	1190	73
	Ind. 2	574	248	381	576	545	380	440	27
25	Ind. 1	1554	672	1032	1558	1475	1028	1190	74
	Ind. 2	547	236	363	548	519	361	418	26
26	Ind. 1	1554	672	1032	1558	1475	1028	1190	75
	Ind. 2	520	225	345	521	494	344	398	25
27	Ind. 1	1554	672	1032	1558	1475	1028	1190	76
	Ind. 2	489	211	325	490	464	323	374	24
28	Ind. 1	1554	672	1032	1558	1475	1028	1190	77
	Ind. 2	466	201	309	467	442	308	357	23
29	Ind. 1	1554	672	1032	1558	1475	1028	1190	78
	Ind. 2	438	189	291	439	415	289	335	22
30	Ind. 1	1554	672	1032	1558	1475	1028	1190	79
	Ind. 2	413	179	274	414	392	273	316	21
31	Ind. 1	1554	672	1032	1558	1475	1028	1190	80
	Ind. 2	388	168	258	390	369	257	298	20
32	Ind. 1	1554	672	1032	1558	1475	1028	1190	91
	Ind. 2	155	67	103	155	147	102	119	9

Tab. 9 RFU pair values arrays of Individuals 1 and 2.

5.4 Construction of the universal matrix

Due to space constraints, it is impossible to exhaustively list the 32 mock mixtures produced in this study; for simplicity, seven of them are partially listed in Appendix A (Page 83). A representative image of each mock mixture is also provided for full reference.

5.5 Statistical analysis

The LRs reported in Table 10 were obtained from the 32 groups of data:

		DYS576	DYS389II	DYS627	DYS458	DYS19	DYS448	DYS391		
		Haplotypes								
Ind.1 22-028-A		19	30	23	18	14	19	12		
Ind.2 22-028-B		20	29	21	15	17	20	10		
Mock Mixture		Peak Height (RFU)							Amount (%)	LR
1	Ind. 1	1554	672	1032	1558	1475	1028	1190	50,3	81,1609
	Ind. 2	1538	665	1021	1542	1460	1017	1178	49,7	
2	Ind. 1	1554	672	1032	1558	1475	1028	1190	51	92,4829
	Ind. 2	1493	645	991	1497	1417	987	1143	49	
3	Ind. 1	1554	672	1032	1558	1475	1028	1190	52	115,6592
	Ind. 2	1434	620	952	1438	1362	949	1098	48	
4	Ind. 1	1554	672	1032	1558	1475	1028	1190	53	135,0530
	Ind. 2	1398	604	928	1402	1327	925	1071	47	
5	Ind. 1	1554	672	1032	1558	1475	1028	1190	54	195,7021
	Ind. 2	1320	571	877	1324	1253	873	1011	46	
6	Ind. 1	1554	672	1032	1558	1475	1028	1190	55	258,8065
	Ind. 2	1271	550	844	1275	1207	841	974	45	
7	Ind. 1	1554	672	1032	1558	1475	1028	1190	56	352,5754
	Ind. 2	1221	528	811	1224	1159	808	935	44	
8	Ind. 1	1554	672	1032	1558	1475	1028	1190	57	516,6339
	Ind. 2	1165	504	774	1168	1106	771	892	43	
9	Ind. 1	1554	672	1032	1558	1475	1028	1190	58	694,5889
	Ind. 2	1125	486	747	1127	1067	744	861	42	
10	Ind. 1	1554	672	1032	1558	1475	1028	1190	59	953,8290
	Ind. 2	1079	466	716	1082	1024	713	826	41	
11	Ind. 1	1554	672	1032	1558	1475	1028	1190	60	1.308,3760
	Ind. 2	1035	447	687	1038	983	685	793	40	
12	Ind. 1	1554	672	1032	1558	1475	1028	1190	61	1.888,8304
	Ind. 2	993	429	659	995	942	656	760	39	
13	Ind. 1	1554	672	1032	1558	1475	1028	1190	62	2.213,9267
	Ind. 2	1041	409	628	948	898	626	724	38	
14	Ind. 1	1554	672	1032	1558	1475	1028	1190	63	3.052,1789
	Ind. 2	916	396	608	919	870	606	702	37	
15	Ind. 1	1554	672	1032	1558	1475	1028	1190	64	3.519,1867
	Ind. 2	878	379	583	880	833	580	672	36	
16	Ind. 1	1554	672	1032	1558	1475	1028	1190	65	3.991,5344
	Ind. 2	834	360	554	836	792	552	639	35	
17	Ind. 1	1554	672	1032	1558	1475	1028	1190	66	4.284,8827
	Ind. 2	798	345	530	800	758	528	611	34	
18	Ind. 1	1554	672	1032	1558	1475	1028	1190	67	4.506,5572
	Ind. 2	761	329	505	763	722	503	583	33	
19	Ind. 1	1554	672	1032	1558	1475	1028	1190	68	4.639,0717
	Ind. 2	731	316	485	733	694	483	560	32	
20	Ind. 1	1554	672	1032	1558	1475	1028	1190	69	4.761,9465
	Ind. 2	696	301	462	697	660	460	533	31	
21	Ind. 1	1554	672	1032	1558	1475	1028	1190	70	4.852,4270
	Ind. 2	665	287	441	666	631	439	509	30	
22	Ind. 1	1554	672	1032	1558	1475	1028	1190	71	4.917,1602
	Ind. 2	637	275	423	638	604	421	487	29	
23	Ind. 1	1554	672	1032	1558	1475	1028	1190	72	4.957,9333
	Ind. 2	606	262	402	607	575	400	464	28	
24	Ind. 1	1554	672	1032	1558	1475	1028	1190	73	4.979,6970
	Ind. 2	574	248	381	576	545	380	440	27	
25	Ind. 1	1554	672	1032	1558	1475	1028	1190	74	4.989,7402
	Ind. 2	547	236	363	548	519	361	418	26	
26	Ind. 1	1554	672	1032	1558	1475	1028	1190	75	4.994,8336
	Ind. 2	520	225	345	521	494	344	398	25	
27	Ind. 1	1554	672	1032	1558	1475	1028	1190	76	4.997,8107
	Ind. 2	489	211	325	490	464	323	374	24	
28	Ind. 1	1554	672	1032	1558	1475	1028	1190	77	4.998,8783
	Ind. 2	466	201	309	467	442	308	357	23	
29	Ind. 1	1554	672	1032	1558	1475	1028	1190	78	4.999,5290
	Ind. 2	438	189	291	439	415	289	335	22	
30	Ind. 1	1554	672	1032	1558	1475	1028	1190	79	4.999,7908
	Ind. 2	413	179	274	414	392	273	316	21	
31	Ind. 1	1554	672	1032	1558	1475	1028	1190	80	4.999,9103
	Ind. 2	388	168	258	390	369	257	298	20	
32	Ind. 1	1554	672	1032	1558	1475	1028	1190	91	5.000,0000
	Ind. 2	155	67	103	155	147	102	119	9	

Tab. 10 LRs result for each mock mixture.

Through the developed calculation system, the aim was to identify whether the haplotype of interest is recognized in the mixture, and for this, it was necessary to consider three cardinal principles:

- i. Under conditions of quantitative haplotype equilibrium, all combinatorial solutions are equiprobable, and there is uncertainty about the presence/absence of POIs as contributors within the mixed evidence;
- ii. Mixtures with a quantitative imbalance between contributors are easily interpretable;
- iii. The WoE extractable from mixtures of the latter kind tends to the value of $1/\text{frequency}$ of the haplotype individually considered.

These considerations regarding the influence of quantitative analysis on the LR can be easily appreciated from the graph in Figure 18, in which it is evident how the imbalance in quantity between haplotype 1 and haplotype 2 causes haplotype 1 to gain probability of existence until it reaches a maximum point at which the probability of haplotype 1 being in the mixture flattens out with its probability of existence in the population at large.

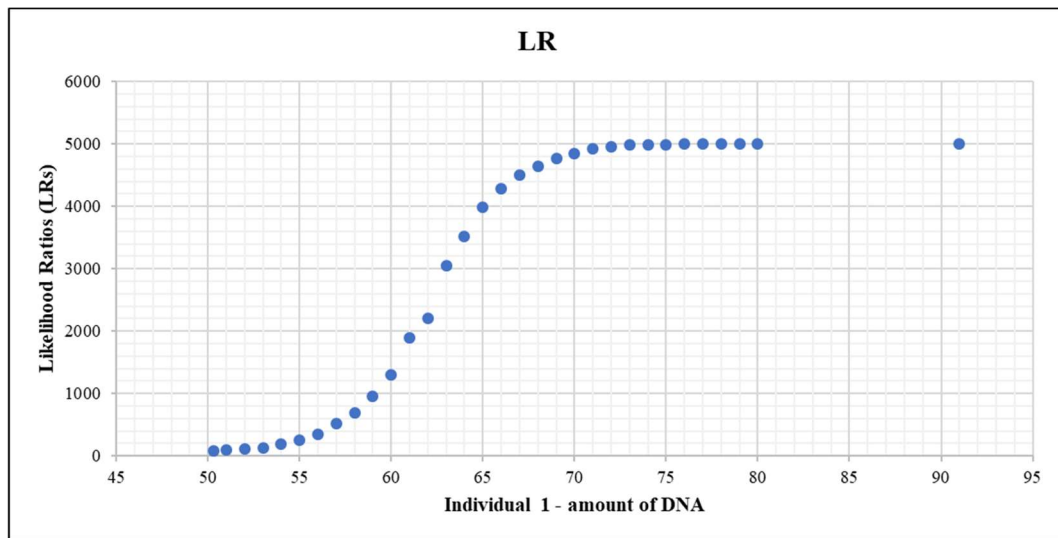


Fig. 18 Likelihood Ratio results for the 32 mock mixtures.

DISCUSSION

We have introduced an original procedure aimed at calculating likelihood ratios to define the probability of a given Person of Interest being present within a mixture typed by Y-Chr profiles that is supposed to contain two persons. Our procedure also enables a complete deconvolution of Y-Chr two-person mixtures, listing all possible combinations (in this case called permutations) resulting from the permutation process.

Both instances were addressed to provide full Bayesian quantitative evidence: deconvolution and LR calculations referring to a particular POI.

Extracting evidence from a Y-Chr mixture is a relatively unexplored issue of modern forensic genetics, especially compared to the case of autosomal mixtures, which has attracted the attention of many researchers in the last few decades.

Papers dealing with the particular issue of Y-Chr mixtures are sporadic, and software-enabling calculations of this kind are nearly completely missing.

Wolf et al. pioneered the field in 2005. Their study exploited population frequency information and introduced two approaches involving the inclusion/exclusion principle and a simulation based on the recursion of the number of unknown contributors [Wolf et al., 2005].

More recently, Professor David Balding and his research group have explored the case of unbalanced Y-Chr mixtures using a simulation approach and reported that in 2PM and possibly 4PM, mixed evidence (E_m) does not result in a significant loss of evidentiary power compared to the single-source evidence (E_{ss}), provided that the E_m contains a predominant profile [Andersen & Balding, 2019b]. This is somewhat an unexpected finding that softens a widely held traditional assumption in the field of mixture analysis:

$$WoE_{ss} > WoE_m$$

In other words, the weight of evidence is considerably weakened while shifting from the single source to the mixed evidence.

Within the simulation framework provided by Andersen and Balding, a POI profile included in mixed evidence with up to four contributors returns almost as strong

evidence as the single-contributor POI profile. In particular, despite being included in a 2-male mixture, the POI profile yields virtually the same evidence value as a single-contributor match. This concept stems from the observation that although millions of possible haplotype profiles could have generated the mixture, very few are likely to exist in the population. Consequently, according to Andersen and Balding, a 2-male mixed evidence profile can be presented to the Court as the single-contributor profile counterpart [Andersen & Balding, 2019b]. In other words, when there is an unbalanced mixture with a major profile, the latter can be reported as being present individually.

Excluding the case of these two groups of Authors, we found no other significant contributions to the theory of haploid mixture interpretation in the current Literature. This not only holds for the very complex case of Bayesian quantitative interpretation but – somewhat surprisingly – for the qualitative semicontinuous approach, which, in principle, should be much simpler to model and use routinely. In the almost complete absence of a reference theory, expert testimony in Court in cases of mixed Y-Chr traces has long been a source of possibly reckless empiricism. Within this background, the soundest way adopted to report in Court has been the *Random Man not Excluded* approach.

This situation has motivated us to tackle the issue and devise a scheme of calculation of our conception.

Our approach starts with drawing an exact map of all possible haplotypes to be generated by letting genetic marker variants permutate across X contributors. This is, in principle, an X^n matrix generating the permutations represented in Table 11:

		N. of genetic markers (n)		
		7	15	24
N. of contributors (X)	2	128	32768	1,68E+07
	3	2187	1,43E+07	2,82E+11
	4	16384	1,07E+09	2,81E+14

Tab. 11 Reconstruction of the number of possible permutations in an X^n matrix. In yellow are highlighted the number of permutations that can be accommodated within a standard Excel worksheet, in grey are all those that cannot

It follows that an ordinary computer can be used to build 2PM matrices encompassing the minimal set of seven haplotype markers up to a set of 10 markers. Therefore, we focused on the 2PM scheme. As we shall see, this model will allow us to demonstrate that our statistical approach works well by building the relevant permutations within the Excel environment and exploiting its typical facilities. A typical example of 10 markers for a 2-person matrix is shown in Figures 10 and 11 (Pages 53-54).

The two-person mixture offers several advantages to all those who aim to model the mixed Y-Chr evidence.

First, the complete deconvolution of up to 10 loci haplotypes is made of a reasonable number of records that can still be accommodated into a standard worksheet. Increasing the number of loci or contributors will give rise to enormous matrices/deconvolution states that electronic spreadsheets cannot usually handle. Even worse, the computational capacity of all commercially available personal computers will not be enough to bear all the calculations required to interpret this kind of mixture.

Second, the calculation of a likelihood involving a known Person of Interest is greatly facilitated by the fact that this case reduces to just one record in the matrix – multiple records will denote the case of likelihoods of three/four-person mixtures. Third, the 2PM case was addressed to model deconvolutions involving evidence consisting of two peaks only (the one peak case is of no interest because it yields

LR that invariably amounts to one). Things become much more complicated when it comes to dealing with 3PM, whose proper evidence is made of one peak, two peak and three peak cases (the latter two being those that can be useful), and with 4PM (involving three kinds of helpful evidence; the two-, the three- and the four-peak cases).

In this paper, we will deal with calculations involving the case of only seven loci/two peaks applied to the 2-person mixture.

Although this might seem a somewhat limited approach – compared to the huge typing power of modern kits/technologies based on dozens of markers – it will become clear that this is not a meaningful limit. Two-person mixtures expressing only a handful of 2-peak loci (actually, seven) often arise from the adoption of kits composed of several loci (typically 15 to 24), expressing several useless EPGs made of only one peak per locus (two persons sharing the same peak). Moreover, it will become apparent that the quantitative approach based on just seven loci may, in many cases, express very meaningful LRs and enable the settlement of whatever issue of POI identification in a haploid mixture.

The first step in making computations in the matter of Y-Chr – 2PM is building the relevant matrices. A matrix is essentially the complete deconvolution of a mixture. If using a universal nomenclature, the 7loci – 2PM matrix becomes the general deconvolution of this class of mixtures to be used in every case. This is somewhat of an original finding, as many Authors have repeatedly hinted at the vast body of permuted haplotypes but never reported the relevant repertory.

Matrices of this sort are the backbone for all subsequent calculations. They are also a typical and very effective way to report the deconvolution of any given mixed Y-Chr evidence.

A simple way to use haplotype matrices and draw Bayesian inference from a 2PM trace is to give a population frequency to all haplotypes generated by the process. This is, in fact, a semi-continuous approach.

If we have a POI to work with, the relevant LR end up to contain the POI haplotype frequency (HF_{POI}) at the numerator and the sum of haplotype frequencies of all combinatorically possible haplotypes at the denominator ($\sum X^n HF$):

$$LR = \frac{HF_{POI}}{\sum X^n HF}$$

which leaves the operator to cope with the impossible task of retrieving hundreds, thousands and possibly millions of theoretical haplotypes within databases (such as the YHRD [Willuweit & Roewer, 2015] that contain 300.000 of them at most. Unsurprisingly, most of these haplotypes do not exist in reality, and their frequency is zero. If bringing this scenario to its extreme consequences and assuming that no other haplotypes than the POIs one are retrieved in the database, then the LR mentioned above reduces to

$$LR = \frac{1}{HF_{POI}}$$

This is a finding that corroborates Andersen and Blading's claims [Andersen & Balding, 2019b].

The weak point of this reasoning is that we cannot ensure that all existing Y-Chr databases will reflect the reality of the real world of human lineages in the general population.

This gives a clear idea of how frustrating the task of calculating semi-continuous likelihoods in the field of Y-Chr mixtures can be.

While reckoning on this, we decided to go further and switch to the quantitative side of likelihood calculations.

The rationale behind the quantitative approach resides in the assumption that the peak heights/peak areas are distributed across loci in the EPGs so that the analyst can always discriminate what peaks belong to either contributor.

The quantitative ratio among two contributors participating in a mix (i.e. the MR) is an obvious parameter to evaluate this consistency. The MR is a universal key for

interpreting all kinds of mixed traces in forensics, and this well-known principle can also be applied to the Y-Chr profiles.

Another valuable principle to adopt is internal coherence, which is applied to quantitative data originating from the same contributor. A way to explain this principle is: provided that the n^{th} contributor has contributed η cells, then η molecules must be retrieved within each locus of his profile. Deviations from this amount may also exist due to our inability to count the amount of molecules in an exact way. This is the only source of uncertainty in the within-locus coherent number of molecules, and it must be accounted for.

We devised statistical indexes to account for, in turn, the mix ratio across loci and the quantitative coherence within loci of the same contributor. The best proxies for the two concepts are the $(1-\chi^2)$ index and the within-contributor PhR.

Chapter 4, Materials and Methods, introduces and describes these two indexes.

At a seven-markers/two-person mixture, there can be 896 $(1-\chi^2)$ calculations ($7 \times 128 = 896$) and 5376 PhR calculations ($42 \times 128 = 5376$).

The number of these calculations increases with increasing number of loci. Calculations are such as to increase the probability of the true permutation and selectively decrease the probability of all others. So, the resulting LR is expected to increase as long as genetic markers are added to the typing protocol.

Both calculation indexes converge towards the same value of likelihood so that the entire value of the evidence to be expressed by the quantitative evidence outweighs that contributed by population probabilities by several orders of magnitude. This is equivalent to saying that one can neglect population haplotypes when it comes to formulating a realistic probability for a given POI to be in the evidence.

This is one of the best features of our methodology.

Another essential property of our quantitative evidence is that its value increases as long as the mixture becomes unbalanced. It follows that well-balanced mixtures are complicated to solve univocally because all states of deconvolutions are as equally plausible according to the quantitative criterion. No real mixture is really perfectly balanced – and to some extent, every mixture should give, sooner or later, the analyst a chance to solve. Of course, well-balanced mixtures will need to pie up an

exceeding number of loci to give the right combination a sufficient chance to become noticed as the most likely one.

Dealing with quantities somewhat reclaims the power of individualization lost by a single source profile while merging with another. Under this perspective, the notation reported on Page 72 should be generalized into

$$WoE_{ss} \geq WoE_m$$

However, the driving force equalizing the two values at an unbalanced mixture is the quantitative factor and not the gene frequencies factor.

It did not escape to us that computing on the quantitative side could possibly have a profound impact on the case of mixed Y evidence because it could lead to minimizing the impact and importance of building big databases and using them for real-world computations.

So far, we do not have a model to apply to cases of haplotype exclusions. This should be cited as one of the main limitations of our work.

Haplotypes of POIs that are not among the true contributors to a mixture may differ from the E_m by a partial share of alleles within a mainframe of compatible loci. In this paper, we manage to compute at compatible loci effectively, but we are still searching for a way to apply quantities and yield below zero LR in incompatible cases.

We will reserve to explore this residual – but crucial – case of the mixed Y chromosome evidence in the near future.

REFERENCES

- Andersen, M. M., & Balding, D. J. (2019b). Y-profile evidence: Close paternal relatives and mixtures. *Forensic Science International: Genetics*, 38, 48–53. <https://doi.org/10.1016/J.FSIGEN.2018.10.004>
- Willuweit, S., & Roewer, L. (2015). The new Y Chromosome Haplotype Reference Database. *Forensic Science International. Genetics*, 15, 43–48. <https://doi.org/10.1016/J.FSIGEN.2014.11.024>
- Wolf, A., Caliebe, A., Junge, O., & Krawczak, M. (2005). Forensic interpretation of Y-chromosomal DNA mixtures. *Forensic Science International*, 152(2–3), 209–213. <https://doi.org/10.1016/j.forsciint.2004.07.021>

CHAPTER 6

GENERAL CONCLUSIONS..... 81

GENERAL CONCLUSIONS

The deconvolution of Y DNA mixtures represents a relatively unexplored area of research posing significant challenges. We sought to develop a computational prototype enabling us to estimate the evidential value of a biological mixture under the assumption that it can come from two individuals.

Following a critical analysis of the literature available on the subject, the Y-STR haplotypes of several male donors were determined, with the view of selecting two individual haplotypes that shared the smallest possible number of alleles. In the initial phase, ten markers on the Y chromosome were considered to test the robustness of our theoretical assumption. However, due to the computational bottleneck involved by the commercial PC processor available, the number of markers was subsequently reduced to seven.

A Microsoft Excel spreadsheet was initially developed to generate a matrix containing all states of permutation, i.e., all possible haplotype combinations, based on the number of loci and the number of contributors involved. This tool is available upon direct request to the Authors.

Once the matrix construction was defined, three independent indices were identified and calculated:

- the haplotypic frequency of the person of interest (POI);
- the peak ratio within each locus;
- the peak ratio across different loci of the mixed stain.

These parameters were implemented in a second Excel spreadsheet, which automatically calculates the likelihood ratio (LR) based on the generated matrix. This tool is also available upon direct request to the Authors.

To the best of our knowledge, our calculation prototype introduces, for the first time, an approach based on a continuous Bayesian model for the analysis of simple Y-STR DNA mixtures.

We obtained promising results, especially in the case of highly unbalanced mixtures where individual contributors are easier to identify.

Future directions of our work will envision: a) a comprehensive study of the proposed model, with particular attention to the effects of stutter, which could compromise the correct interpretation of haplotypes, especially in the case of

contiguous alleles; b) the application of the model to real cases to assess its effectiveness on mixed stains recovered from crime scenes and analyzed using gonosomal markers.

Appendix A

Mixture 0.503 – 0.497

	Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	Ind. 1 Peak Height (RFU)				Ind. 2 Peak Height (RFU)									
1	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	19,30,23,18,14,19,12	1538,665,1021,1542,1460,1017,1178	1554	672	1032	1558	1475	1028	1190	1538	665	1021	1542	1460	1017	1178
2	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,1178	19,30,23,18,14,19,10	1538,665,1021,1542,1460,1017,1190	1554	672	1032	1558	1475	1028	1178	1538	665	1021	1542	1460	1017	1190
3	20,29,21,15,17,19,10	1554,672,1032,1558,1475,1017,1190	19,30,23,18,14,20,12	1538,665,1021,1542,1460,1028,1178	1554	672	1032	1558	1475	1017	1190	1538	665	1021	1542	1460	1028	1178
4	20,29,21,15,17,19,12	1554,672,1032,1558,1475,1017,1178	19,30,23,18,14,20,10	1538,665,1021,1542,1460,1028,1190	1554	672	1032	1558	1475	1017	1178	1538	665	1021	1542	1460	1028	1190
5	20,29,21,15,14,20,10	1554,672,1032,1558,1460,1028,1190	19,30,23,18,17,19,12	1538,665,1021,1542,1475,1017,1178	1554	672	1032	1558	1460	1028	1190	1538	665	1021	1542	1475	1017	1178
6	20,29,21,15,14,20,12	1554,672,1032,1558,1460,1028,1178	19,30,23,18,17,19,10	1538,665,1021,1542,1475,1017,1190	1554	672	1032	1558	1460	1028	1178	1538	665	1021	1542	1475	1017	1190
7	20,29,21,15,14,19,10	1554,672,1032,1558,1460,1017,1190	19,30,23,18,17,20,12	1538,665,1021,1542,1475,1028,1178	1554	672	1032	1558	1460	1017	1190	1538	665	1021	1542	1475	1028	1178
8	20,29,21,15,14,19,12	1554,672,1032,1558,1460,1017,1178	19,30,23,18,17,20,10	1538,665,1021,1542,1475,1028,1190	1554	672	1032	1558	1460	1017	1178	1538	665	1021	1542	1475	1028	1190
9	20,29,21,18,17,20,10	1554,672,1032,1542,1475,1028,1190	19,30,23,15,14,19,12	1538,665,1021,1558,1460,1017,1178	1554	672	1032	1542	1475	1028	1190	1538	665	1021	1558	1460	1017	1178
10	20,29,21,18,17,20,12	1554,672,1032,1542,1475,1028,1178	19,30,23,15,14,19,10	1538,665,1021,1558,1460,1017,1190	1554	672	1032	1542	1475	1028	1178	1538	665	1021	1558	1460	1017	1190
			...															
117	19,30,23,15,14,20,10	1538,665,1021,1558,1460,1028,1190	20,29,21,18,17,19,12	1554,672,1032,1542,1475,1017,1178	1538	665	1021	1558	1460	1028	1190	1554	672	1032	1542	1475	1017	1178
118	19,30,23,15,14,20,12	1538,665,1021,1558,1460,1028,1178	20,29,21,18,17,19,10	1554,672,1032,1542,1475,1017,1190	1538	665	1021	1558	1460	1028	1178	1554	672	1032	1542	1475	1017	1190
119	19,30,23,15,14,19,10	1538,665,1021,1558,1460,1017,1190	20,29,21,18,17,20,12	1554,672,1032,1542,1475,1028,1178	1538	665	1021	1558	1460	1017	1190	1554	672	1032	1542	1475	1028	1178
120	19,30,23,15,14,19,12	1538,665,1021,1558,1460,1017,1178	20,29,21,18,17,20,10	1554,672,1032,1542,1475,1028,1190	1538	665	1021	1558	1460	1017	1178	1554	672	1032	1542	1475	1028	1190
121	19,30,23,18,17,20,10	1538,665,1021,1542,1475,1028,1190	20,29,21,15,14,19,12	1554,672,1032,1558,1460,1017,1178	1538	665	1021	1542	1475	1028	1190	1554	672	1032	1558	1460	1017	1178
122	19,30,23,18,17,20,12	1538,665,1021,1542,1475,1028,1178	20,29,21,15,14,19,10	1554,672,1032,1558,1460,1017,1190	1538	665	1021	1542	1475	1028	1178	1554	672	1032	1558	1460	1017	1190
123	19,30,23,18,17,19,10	1538,665,1021,1542,1475,1017,1190	20,29,21,15,14,20,12	1554,672,1032,1558,1460,1028,1178	1538	665	1021	1542	1475	1017	1190	1554	672	1032	1558	1460	1028	1178
124	19,30,23,18,17,19,12	1538,665,1021,1542,1475,1017,1178	20,29,21,15,14,20,10	1554,672,1032,1558,1460,1028,1190	1538	665	1021	1542	1475	1017	1178	1554	672	1032	1558	1460	1028	1190
125	19,30,23,18,14,20,10	1538,665,1021,1542,1460,1028,1190	20,29,21,15,17,19,12	1554,672,1032,1558,1475,1017,1178	1538	665	1021	1542	1460	1028	1190	1554	672	1032	1558	1475	1017	1178
126	19,30,23,18,14,20,12	1538,665,1021,1542,1460,1028,1178	20,29,21,15,17,19,10	1554,672,1032,1558,1475,1017,1190	1538	665	1021	1542	1460	1028	1178	1554	672	1032	1558	1475	1017	1190
127	19,30,23,18,14,19,10	1538,665,1021,1542,1460,1017,1190	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,1178	1538	665	1021	1542	1460	1017	1190	1554	672	1032	1558	1475	1028	1178
128	19,30,23,18,14,19,12	1538,665,1021,1542,1460,1017,1178	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	1538	665	1021	1542	1460	1017	1178	1554	672	1032	1558	1475	1028	1190

Mixture 0.55 – 0.45

	Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	Ind. 1 Peak Height (RFU)				Ind. 2 Peak Height (RFU)									
1	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	19,30,23,18,14,19,12	1271,550,844,1275,1207,841,974	1554	672	1032	1558	1475	1028	1190	1271	550	844	1275	1207	841	974
2	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,974	19,30,23,18,14,19,10	1271,550,844,1275,1207,841,1190	1554	672	1032	1558	1475	1028	974	1271	550	844	1275	1207	841	1190
3	20,29,21,15,17,19,10	1554,672,1032,1558,1475,841,1190	19,30,23,18,14,20,12	1271,550,844,1275,1207,1028,974	1554	672	1032	1558	1475	841	1190	1271	550	844	1275	1207	1028	974
4	20,29,21,15,17,19,12	1554,672,1032,1558,1475,841,974	19,30,23,18,14,20,10	1271,550,844,1275,1207,1028,1190	1554	672	1032	1558	1475	841	974	1271	550	844	1275	1207	1028	1190
5	20,29,21,15,14,20,10	1554,672,1032,1558,1207,1028,1190	19,30,23,18,17,19,12	1271,550,844,1275,1475,841,974	1554	672	1032	1558	1207	1028	1190	1271	550	844	1275	1475	841	974
6	20,29,21,15,14,20,12	1554,672,1032,1558,1207,1028,974	19,30,23,18,17,19,10	1271,550,844,1275,1475,841,1190	1554	672	1032	1558	1207	1028	974	1271	550	844	1275	1475	841	1190
7	20,29,21,15,14,19,10	1554,672,1032,1558,1207,841,1190	19,30,23,18,17,20,12	1271,550,844,1275,1475,1028,974	1554	672	1032	1558	1207	841	1190	1271	550	844	1275	1475	1028	974
8	20,29,21,15,14,19,12	1554,672,1032,1558,1207,841,974	19,30,23,18,17,20,10	1271,550,844,1275,1475,1028,1190	1554	672	1032	1558	1207	841	974	1271	550	844	1275	1475	1028	1190
9	20,29,21,18,17,20,10	1554,672,1032,1275,1475,1028,1190	19,30,23,15,14,19,12	1271,550,844,1558,1207,841,974	1554	672	1032	1275	1475	1028	1190	1271	550	844	1558	1207	841	974
10	20,29,21,18,17,20,12	1554,672,1032,1275,1475,1028,974	19,30,23,15,14,19,10	1271,550,844,1558,1207,841,1190	1554	672	1032	1275	1475	1028	974	1271	550	844	1558	1207	841	1190
			...															
117	19,30,23,15,14,20,10	1271,550,844,1558,1207,1028,1190	20,29,21,18,17,19,12	1554,672,1032,1275,1475,841,974	1271	550	844	1558	1207	1028	1190	1554	672	1032	1275	1475	841	974
118	19,30,23,15,14,20,12	1271,550,844,1558,1207,1028,974	20,29,21,18,17,19,10	1554,672,1032,1275,1475,841,1190	1271	550	844	1558	1207	1028	974	1554	672	1032	1275	1475	841	1190
119	19,30,23,15,14,19,10	1271,550,844,1558,1207,841,1190	20,29,21,18,17,20,12	1554,672,1032,1275,1475,1028,974	1271	550	844	1558	1207	841	1190	1554	672	1032	1275	1475	1028	974
120	19,30,23,15,14,19,12	1271,550,844,1558,1207,841,974	20,29,21,18,17,20,10	1554,672,1032,1275,1475,1028,1190	1271	550	844	1558	1207	841	974	1554	672	1032	1275	1475	1028	1190
121	19,30,23,18,17,20,10	1271,550,844,1275,1475,1028,1190	20,29,21,15,14,19,12	1554,672,1032,1558,1207,841,974	1271	550	844	1275	1475	1028	1190	1554	672	1032	1558	1207	841	974
122	19,30,23,18,17,20,12	1271,550,844,1275,1475,1028,974	20,29,21,15,14,19,10	1554,672,1032,1558,1207,841,1190	1271	550	844	1275	1475	1028	974	1554	672	1032	1558	1207	841	1190
123	19,30,23,18,17,19,10	1271,550,844,1275,1475,841,1190	20,29,21,15,14,20,12	1554,672,1032,1558,1207,1028,974	1271	550	844	1275	1475	841	1190	1554	672	1032	1558	1207	1028	974
124	19,30,23,18,17,19,12	1271,550,844,1275,1475,841,974	20,29,21,15,14,20,10	1554,672,1032,1558,1207,1028,1190	1271	550	844	1275	1475	841	974	1554	672	1032	1558	1207	1028	1190
125	19,30,23,18,14,20,10	1271,550,844,1275,1207,1028,1190	20,29,21,15,17,19,12	1554,672,1032,1558,1475,841,974	1271	550	844	1275	1207	1028	1190	1554	672	1032	1558	1475	841	974
126	19,30,23,18,14,20,12	1271,550,844,1275,1207,1028,974	20,29,21,15,17,19,10	1554,672,1032,1558,1475,841,1190	1271	550	844	1275	1207	1028	974	1554	672	1032	1558	1475	841	1190
127	19,30,23,18,14,19,10	1271,550,844,1275,1207,841,1190	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,974	1271	550	844	1275	1207	841	1190	1554	672	1032	1558	1475	1028	974
128	19,30,23,18,14,19,12	1271,550,844,1275,1207,841,974	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	1271	550	844	1275	1207	841	974	1554	672	1032	1558	1475	1028	1190

Mixture 0.60 – 0.40

	Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	Ind. 1 Peak Height (RFU)				Ind. 2 Peak Height (RFU)									
1	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	19,30,23,18,14,19,12	1035,447,687,1038,983,685,793	1554	672	1032	1558	1475	1028	1190	1035	447	687	1038	983	685	793
2	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,793	19,30,23,18,14,19,10	1035,447,687,1038,983,685,1190	1554	672	1032	1558	1475	1028	793	1035	447	687	1038	983	685	1190
3	20,29,21,15,17,19,10	1554,672,1032,1558,1475,685,1190	19,30,23,18,14,20,12	1035,447,687,1038,983,1028,793	1554	672	1032	1558	1475	685	1190	1035	447	687	1038	983	1028	793
4	20,29,21,15,17,19,12	1554,672,1032,1558,1475,685,793	19,30,23,18,14,20,10	1035,447,687,1038,983,1028,1190	1554	672	1032	1558	1475	685	793	1035	447	687	1038	983	1028	1190
5	20,29,21,15,14,20,10	1554,672,1032,1558,983,1028,1190	19,30,23,18,17,19,12	1035,447,687,1038,1475,685,793	1554	672	1032	1558	983	1028	1190	1035	447	687	1038	1475	685	793
6	20,29,21,15,14,20,12	1554,672,1032,1558,983,1028,793	19,30,23,18,17,19,10	1035,447,687,1038,1475,685,1190	1554	672	1032	1558	983	1028	793	1035	447	687	1038	1475	685	1190
7	20,29,21,15,14,19,10	1554,672,1032,1558,983,685,1190	19,30,23,18,17,20,12	1035,447,687,1038,1475,1028,793	1554	672	1032	1558	983	685	1190	1035	447	687	1038	1475	1028	793
8	20,29,21,15,14,19,12	1554,672,1032,1558,983,685,793	19,30,23,18,17,20,10	1035,447,687,1038,1475,1028,1190	1554	672	1032	1558	983	685	793	1035	447	687	1038	1475	1028	1190
9	20,29,21,18,17,20,10	1554,672,1032,1038,1475,1028,1190	19,30,23,15,14,19,12	1035,447,687,1558,983,685,793	1554	672	1032	1038	1475	1028	1190	1035	447	687	1558	983	685	793
10	20,29,21,18,17,20,12	1554,672,1032,1038,1475,1028,793	19,30,23,15,14,19,10	1035,447,687,1558,983,685,1190	1554	672	1032	1038	1475	1028	793	1035	447	687	1558	983	685	1190
			...															
117	19,30,23,15,14,20,10	1035,447,687,1558,983,1028,1190	20,29,21,18,17,19,12	1554,672,1032,1038,1475,685,793	1035	447	687	1558	983	1028	1190	1554	672	1032	1038	1475	685	793
118	19,30,23,15,14,20,12	1035,447,687,1558,983,1028,793	20,29,21,18,17,19,10	1554,672,1032,1038,1475,685,1190	1035	447	687	1558	983	1028	793	1554	672	1032	1038	1475	685	1190
119	19,30,23,15,14,19,10	1035,447,687,1558,983,685,1190	20,29,21,18,17,20,12	1554,672,1032,1038,1475,1028,793	1035	447	687	1558	983	685	1190	1554	672	1032	1038	1475	1028	793
120	19,30,23,15,14,19,12	1035,447,687,1558,983,685,793	20,29,21,18,17,20,10	1554,672,1032,1038,1475,1028,1190	1035	447	687	1558	983	685	793	1554	672	1032	1038	1475	1028	1190
121	19,30,23,18,17,20,10	1035,447,687,1038,1475,1028,1190	20,29,21,15,14,19,12	1554,672,1032,1558,983,685,793	1035	447	687	1038	1475	1028	1190	1554	672	1032	1558	983	685	793
122	19,30,23,18,17,20,12	1035,447,687,1038,1475,1028,793	20,29,21,15,14,19,10	1554,672,1032,1558,983,685,1190	1035	447	687	1038	1475	1028	793	1554	672	1032	1558	983	685	1190
123	19,30,23,18,17,19,10	1035,447,687,1038,1475,685,1190	20,29,21,15,14,20,12	1554,672,1032,1558,983,1028,793	1035	447	687	1038	1475	685	1190	1554	672	1032	1558	983	1028	793
124	19,30,23,18,17,19,12	1035,447,687,1038,1475,685,793	20,29,21,15,14,20,10	1554,672,1032,1558,983,1028,1190	1035	447	687	1038	1475	685	793	1554	672	1032	1558	983	1028	1190
125	19,30,23,18,14,20,10	1035,447,687,1038,983,1028,1190	20,29,21,15,17,19,12	1554,672,1032,1558,1475,685,793	1035	447	687	1038	983	1028	1190	1554	672	1032	1558	1475	685	793
126	19,30,23,18,14,20,12	1035,447,687,1038,983,1028,793	20,29,21,15,17,19,10	1554,672,1032,1558,1475,685,1190	1035	447	687	1038	983	1028	793	1554	672	1032	1558	1475	685	1190
127	19,30,23,18,14,19,10	1035,447,687,1038,983,685,1190	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,793	1035	447	687	1038	983	685	1190	1554	672	1032	1558	1475	1028	793
128	19,30,23,18,14,19,12	1035,447,687,1038,983,685,793	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	1035	447	687	1038	983	685	793	1554	672	1032	1558	1475	1028	1190

Mixture 0.65 – 0.35

	Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	Ind. 1 Peak Height (RFU)								Ind. 2 Peak Height (RFU)							
1	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	19,30,23,18,14,19,12	834,360,554,836,792,552,639	1554	672	1032	1558	1475	1028	1190	834	360	554	836	792	552	639		
2	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,639	19,30,23,18,14,19,10	834,360,554,836,792,552,1190	1554	672	1032	1558	1475	1028	639	834	360	554	836	792	552	1190		
3	20,29,21,15,17,19,10	1554,672,1032,1558,1475,552,1190	19,30,23,18,14,20,12	834,360,554,836,792,1028,639	1554	672	1032	1558	1475	552	1190	834	360	554	836	792	1028	639		
4	20,29,21,15,17,19,12	1554,672,1032,1558,1475,552,639	19,30,23,18,14,20,10	834,360,554,836,792,1028,1190	1554	672	1032	1558	1475	552	639	834	360	554	836	792	1028	1190		
5	20,29,21,15,14,20,10	1554,672,1032,1558,792,1028,1190	19,30,23,18,17,19,12	834,360,554,836,1475,552,639	1554	672	1032	1558	792	1028	1190	834	360	554	836	1475	552	639		
6	20,29,21,15,14,20,12	1554,672,1032,1558,792,1028,639	19,30,23,18,17,19,10	834,360,554,836,1475,552,1190	1554	672	1032	1558	792	1028	639	834	360	554	836	1475	552	1190		
7	20,29,21,15,14,19,10	1554,672,1032,1558,792,552,1190	19,30,23,18,17,20,12	834,360,554,836,1475,1028,639	1554	672	1032	1558	792	552	1190	834	360	554	836	1475	1028	639		
8	20,29,21,15,14,19,12	1554,672,1032,1558,792,552,639	19,30,23,18,17,20,10	834,360,554,836,1475,1028,1190	1554	672	1032	1558	792	552	639	834	360	554	836	1475	1028	1190		
9	20,29,21,18,17,20,10	1554,672,1032,836,1475,1028,1190	19,30,23,15,14,19,12	834,360,554,1558,792,552,639	1554	672	1032	836	1475	1028	1190	834	360	554	1558	792	552	639		
10	20,29,21,18,17,20,12	1554,672,1032,836,1475,1028,639	19,30,23,15,14,19,10	834,360,554,1558,792,552,1190	1554	672	1032	836	1475	1028	639	834	360	554	1558	792	552	1190		
	...																			
117	19,30,23,15,14,20,10	834,360,554,1558,792,1028,1190	20,29,21,18,17,19,12	1554,672,1032,836,1475,552,639	834	360	554	1558	792	1028	1190	1554	672	1032	836	1475	552	639		
118	19,30,23,15,14,20,12	834,360,554,1558,792,1028,639	20,29,21,18,17,19,10	1554,672,1032,836,1475,552,1190	834	360	554	1558	792	1028	639	1554	672	1032	836	1475	552	1190		
119	19,30,23,15,14,19,10	834,360,554,1558,792,552,1190	20,29,21,18,17,20,12	1554,672,1032,836,1475,1028,639	834	360	554	1558	792	552	1190	1554	672	1032	836	1475	1028	639		
120	19,30,23,15,14,19,12	834,360,554,1558,792,552,639	20,29,21,18,17,20,10	1554,672,1032,836,1475,1028,1190	834	360	554	1558	792	552	639	1554	672	1032	836	1475	1028	1190		
121	19,30,23,18,17,20,10	834,360,554,836,1475,1028,1190	20,29,21,15,14,19,12	1554,672,1032,1558,792,552,639	834	360	554	836	1475	1028	1190	1554	672	1032	1558	792	552	639		
122	19,30,23,18,17,20,12	834,360,554,836,1475,1028,639	20,29,21,15,14,19,10	1554,672,1032,1558,792,552,1190	834	360	554	836	1475	1028	639	1554	672	1032	1558	792	552	1190		
123	19,30,23,18,17,19,10	834,360,554,836,1475,552,1190	20,29,21,15,14,20,12	1554,672,1032,1558,792,1028,639	834	360	554	836	1475	552	1190	1554	672	1032	1558	792	1028	639		
124	19,30,23,18,17,19,12	834,360,554,836,1475,552,639	20,29,21,15,14,20,10	1554,672,1032,1558,792,1028,1190	834	360	554	836	1475	552	639	1554	672	1032	1558	792	1028	1190		
125	19,30,23,18,14,20,10	834,360,554,836,792,1028,1190	20,29,21,15,17,19,12	1554,672,1032,1558,1475,552,639	834	360	554	836	792	1028	1190	1554	672	1032	1558	1475	552	639		
126	19,30,23,18,14,20,12	834,360,554,836,792,1028,639	20,29,21,15,17,19,10	1554,672,1032,1558,1475,552,1190	834	360	554	836	792	1028	639	1554	672	1032	1558	1475	552	1190		
127	19,30,23,18,14,19,10	834,360,554,836,792,552,1190	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,639	834	360	554	836	792	552	1190	1554	672	1032	1558	1475	1028	639		
128	19,30,23,18,14,19,12	834,360,554,836,792,552,639	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	834	360	554	836	792	552	639	1554	672	1032	1558	1475	1028	1190		

Mixture 0.70 – 0.30

	Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	Ind. 1 Peak Height (RFU)								Ind. 2 Peak Height (RFU)							
1	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	19,30,23,18,14,19,12	665,287,441,666,631,439,509	1554	672	1032	1558	1475	1028	1190	665	287	441	666	631	439	509		
2	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,509	19,30,23,18,14,19,10	665,287,441,666,631,439,1190	1554	672	1032	1558	1475	1028	509	665	287	441	666	631	439	1190		
3	20,29,21,15,17,19,10	1554,672,1032,1558,1475,439,1190	19,30,23,18,14,20,12	665,287,441,666,631,1028,509	1554	672	1032	1558	1475	439	1190	665	287	441	666	631	1028	509		
4	20,29,21,15,17,19,12	1554,672,1032,1558,1475,439,509	19,30,23,18,14,20,10	665,287,441,666,631,1028,1190	1554	672	1032	1558	1475	439	509	665	287	441	666	631	1028	1190		
5	20,29,21,15,14,20,10	1554,672,1032,1558,631,1028,1190	19,30,23,18,17,19,12	665,287,441,666,1475,439,509	1554	672	1032	1558	631	1028	1190	665	287	441	666	1475	439	509		
6	20,29,21,15,14,20,12	1554,672,1032,1558,631,1028,509	19,30,23,18,17,19,10	665,287,441,666,1475,439,1190	1554	672	1032	1558	631	1028	509	665	287	441	666	1475	439	1190		
7	20,29,21,15,14,19,10	1554,672,1032,1558,631,439,1190	19,30,23,18,17,20,12	665,287,441,666,1475,1028,509	1554	672	1032	1558	631	439	1190	665	287	441	666	1475	1028	509		
8	20,29,21,15,14,19,12	1554,672,1032,1558,631,439,509	19,30,23,18,17,20,10	665,287,441,666,1475,1028,1190	1554	672	1032	1558	631	439	509	665	287	441	666	1475	1028	1190		
9	20,29,21,18,17,20,10	1554,672,1032,666,1475,1028,1190	19,30,23,15,14,19,12	665,287,441,1558,631,439,509	1554	672	1032	666	1475	1028	1190	665	287	441	1558	631	439	509		
10	20,29,21,18,17,20,12	1554,672,1032,666,1475,1028,509	19,30,23,15,14,19,10	665,287,441,1558,631,439,1190	1554	672	1032	666	1475	1028	509	665	287	441	1558	631	439	1190		
				...																
117	19,30,23,15,14,20,10	665,287,441,1558,631,1028,1190	20,29,21,18,17,19,12	1554,672,1032,666,1475,439,509	665	287	441	1558	631	1028	1190	1554	672	1032	666	1475	439	509		
118	19,30,23,15,14,20,12	665,287,441,1558,631,1028,509	20,29,21,18,17,19,10	1554,672,1032,666,1475,439,1190	665	287	441	1558	631	1028	509	1554	672	1032	666	1475	439	1190		
119	19,30,23,15,14,19,10	665,287,441,1558,631,439,1190	20,29,21,18,17,20,12	1554,672,1032,666,1475,1028,509	665	287	441	1558	631	439	1190	1554	672	1032	666	1475	1028	509		
120	19,30,23,15,14,19,12	665,287,441,1558,631,439,509	20,29,21,18,17,20,10	1554,672,1032,666,1475,1028,1190	665	287	441	1558	631	439	509	1554	672	1032	666	1475	1028	1190		
121	19,30,23,18,17,20,10	665,287,441,666,1475,1028,1190	20,29,21,15,14,19,12	1554,672,1032,1558,631,439,509	665	287	441	666	1475	1028	1190	1554	672	1032	1558	631	439	509		
122	19,30,23,18,17,20,12	665,287,441,666,1475,1028,509	20,29,21,15,14,19,10	1554,672,1032,1558,631,439,1190	665	287	441	666	1475	1028	509	1554	672	1032	1558	631	439	1190		
123	19,30,23,18,17,19,10	665,287,441,666,1475,439,1190	20,29,21,15,14,20,12	1554,672,1032,1558,631,1028,509	665	287	441	666	1475	439	1190	1554	672	1032	1558	631	1028	509		
124	19,30,23,18,17,19,12	665,287,441,666,1475,439,509	20,29,21,15,14,20,10	1554,672,1032,1558,631,1028,1190	665	287	441	666	1475	439	509	1554	672	1032	1558	631	1028	1190		
125	19,30,23,18,14,20,10	665,287,441,666,631,1028,1190	20,29,21,15,17,19,12	1554,672,1032,1558,1475,439,509	665	287	441	666	631	1028	1190	1554	672	1032	1558	1475	439	509		
126	19,30,23,18,14,20,12	665,287,441,666,631,1028,509	20,29,21,15,17,19,10	1554,672,1032,1558,1475,439,1190	665	287	441	666	631	1028	509	1554	672	1032	1558	1475	439	1190		
127	19,30,23,18,14,19,10	665,287,441,666,631,439,1190	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,509	665	287	441	666	631	439	1190	1554	672	1032	1558	1475	1028	509		
128	19,30,23,18,14,19,12	665,287,441,666,631,439,509	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	665	287	441	666	631	439	509	1554	672	1032	1558	1475	1028	1190		

Mixture 0.75 – 0.25

	Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	Ind. 1 Peak Height (RFU)							Ind. 2 Peak Height (RFU)						
1	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	19,30,23,18,14,19,12	520,225,345,521,494,344,398	1554	672	1032	1558	1475	1028	1190	520	225	345	521	494	344	398
2	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,398	19,30,23,18,14,19,10	520,225,345,521,494,344,1190	1554	672	1032	1558	1475	1028	398	520	225	345	521	494	344	1190
3	20,29,21,15,17,19,10	1554,672,1032,1558,1475,344,1190	19,30,23,18,14,20,12	520,225,345,521,494,1028,398	1554	672	1032	1558	1475	344	1190	520	225	345	521	494	1028	398
4	20,29,21,15,17,19,12	1554,672,1032,1558,1475,344,398	19,30,23,18,14,20,10	520,225,345,521,494,1028,1190	1554	672	1032	1558	1475	344	398	520	225	345	521	494	1028	1190
5	20,29,21,15,14,20,10	1554,672,1032,1558,494,1028,1190	19,30,23,18,17,19,12	520,225,345,521,1475,344,398	1554	672	1032	1558	494	1028	1190	520	225	345	521	1475	344	398
6	20,29,21,15,14,20,12	1554,672,1032,1558,494,1028,398	19,30,23,18,17,19,10	520,225,345,521,1475,344,1190	1554	672	1032	1558	494	1028	398	520	225	345	521	1475	344	1190
7	20,29,21,15,14,19,10	1554,672,1032,1558,494,344,1190	19,30,23,18,17,20,12	520,225,345,521,1475,1028,398	1554	672	1032	1558	494	344	1190	520	225	345	521	1475	1028	398
8	20,29,21,15,14,19,12	1554,672,1032,1558,494,344,398	19,30,23,18,17,20,10	520,225,345,521,1475,1028,1190	1554	672	1032	1558	494	344	398	520	225	345	521	1475	1028	1190
9	20,29,21,18,17,20,10	1554,672,1032,521,1475,1028,1190	19,30,23,15,14,19,12	520,225,345,1558,494,344,398	1554	672	1032	521	1475	1028	1190	520	225	345	1558	494	344	398
10	20,29,21,18,17,20,12	1554,672,1032,521,1475,1028,398	19,30,23,15,14,19,10	520,225,345,1558,494,344,1190	1554	672	1032	521	1475	1028	398	520	225	345	1558	494	344	1190
	...																	
117	19,30,23,15,14,20,10	520,225,345,1558,494,1028,1190	20,29,21,18,17,19,12	1554,672,1032,521,1475,344,398	520	225	345	1558	494	1028	1190	1554	672	1032	521	1475	344	398
118	19,30,23,15,14,20,12	520,225,345,1558,494,1028,398	20,29,21,18,17,19,10	1554,672,1032,521,1475,344,1190	520	225	345	1558	494	1028	398	1554	672	1032	521	1475	344	1190
119	19,30,23,15,14,19,10	520,225,345,1558,494,344,1190	20,29,21,18,17,20,12	1554,672,1032,521,1475,1028,398	520	225	345	1558	494	344	1190	1554	672	1032	521	1475	1028	398
120	19,30,23,15,14,19,12	520,225,345,1558,494,344,398	20,29,21,18,17,20,10	1554,672,1032,521,1475,1028,1190	520	225	345	1558	494	344	398	1554	672	1032	521	1475	1028	1190
121	19,30,23,18,17,20,10	520,225,345,521,1475,1028,1190	20,29,21,15,14,19,12	1554,672,1032,1558,494,344,398	520	225	345	521	1475	1028	1190	1554	672	1032	1558	494	344	398
122	19,30,23,18,17,20,12	520,225,345,521,1475,1028,398	20,29,21,15,14,19,10	1554,672,1032,1558,494,344,1190	520	225	345	521	1475	1028	398	1554	672	1032	1558	494	344	1190
123	19,30,23,18,17,19,10	520,225,345,521,1475,344,1190	20,29,21,15,14,20,12	1554,672,1032,1558,494,1028,398	520	225	345	521	1475	344	1190	1554	672	1032	1558	494	1028	398
124	19,30,23,18,17,19,12	520,225,345,521,1475,344,398	20,29,21,15,14,20,10	1554,672,1032,1558,494,1028,1190	520	225	345	521	1475	344	398	1554	672	1032	1558	494	1028	1190
125	19,30,23,18,14,20,10	520,225,345,521,494,1028,1190	20,29,21,15,17,19,12	1554,672,1032,1558,1475,344,398	520	225	345	521	494	1028	1190	1554	672	1032	1558	1475	344	398
126	19,30,23,18,14,20,12	520,225,345,521,494,1028,398	20,29,21,15,17,19,10	1554,672,1032,1558,1475,344,1190	520	225	345	521	494	1028	398	1554	672	1032	1558	1475	344	1190
127	19,30,23,18,14,19,10	520,225,345,521,494,344,1190	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,398	520	225	345	521	494	344	1190	1554	672	1032	1558	1475	1028	398
128	19,30,23,18,14,19,12	520,225,345,521,494,344,398	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	520	225	345	521	494	344	398	1554	672	1032	1558	1475	1028	1190

Mixture 0.80 – 0.20

	Ind. 1 haplotype	Ind. 1 Peak Height (RFU)	Ind. 2 haplotype	Ind. 2 Peak Height (RFU)	Ind. 1 Peak Height (RFU)						Ind. 2 Peak Height (RFU)							
1	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	19,30,23,18,14,19,12	388,168,258,390,369,257,298	1554	672	1032	1558	1475	1028	1190	388	168	258	390	369	257	298
2	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,298	19,30,23,18,14,19,10	388,168,258,390,369,257,1190	1554	672	1032	1558	1475	1028	298	388	168	258	390	369	257	1190
3	20,29,21,15,17,19,10	1554,672,1032,1558,1475,257,1190	19,30,23,18,14,20,12	388,168,258,390,369,1028,298	1554	672	1032	1558	1475	257	1190	388	168	258	390	369	1028	298
4	20,29,21,15,17,19,12	1554,672,1032,1558,1475,257,298	19,30,23,18,14,20,10	388,168,258,390,369,1028,1190	1554	672	1032	1558	1475	257	298	388	168	258	390	369	1028	1190
5	20,29,21,15,14,20,10	1554,672,1032,1558,369,1028,1190	19,30,23,18,17,19,12	388,168,258,390,1475,257,298	1554	672	1032	1558	369	1028	1190	388	168	258	390	1475	257	298
6	20,29,21,15,14,20,12	1554,672,1032,1558,369,1028,298	19,30,23,18,17,19,10	388,168,258,390,1475,257,1190	1554	672	1032	1558	369	1028	298	388	168	258	390	1475	257	1190
7	20,29,21,15,14,19,10	1554,672,1032,1558,369,257,1190	19,30,23,18,17,20,12	388,168,258,390,1475,1028,298	1554	672	1032	1558	369	257	1190	388	168	258	390	1475	1028	298
8	20,29,21,15,14,19,12	1554,672,1032,1558,369,257,298	19,30,23,18,17,20,10	388,168,258,390,1475,1028,1190	1554	672	1032	1558	369	257	298	388	168	258	390	1475	1028	1190
9	20,29,21,18,17,20,10	1554,672,1032,390,1475,1028,1190	19,30,23,15,14,19,12	388,168,258,1558,369,257,298	1554	672	1032	390	1475	1028	1190	388	168	258	1558	369	257	298
10	20,29,21,18,17,20,12	1554,672,1032,390,1475,1028,298	19,30,23,15,14,19,10	388,168,258,1558,369,257,1190	1554	672	1032	390	1475	1028	298	388	168	258	1558	369	257	1190
			...															
117	19,30,23,15,14,20,10	388,168,258,1558,369,1028,1190	20,29,21,18,17,19,12	1554,672,1032,390,1475,257,298	388	168	258	1558	369	1028	1190	1554	672	1032	390	1475	257	298
118	19,30,23,15,14,20,12	388,168,258,1558,369,1028,298	20,29,21,18,17,19,10	1554,672,1032,390,1475,257,1190	388	168	258	1558	369	1028	298	1554	672	1032	390	1475	257	1190
119	19,30,23,15,14,19,10	388,168,258,1558,369,257,1190	20,29,21,18,17,20,12	1554,672,1032,390,1475,1028,298	388	168	258	1558	369	257	1190	1554	672	1032	390	1475	1028	298
120	19,30,23,15,14,19,12	388,168,258,1558,369,257,298	20,29,21,18,17,20,10	1554,672,1032,390,1475,1028,1190	388	168	258	1558	369	257	298	1554	672	1032	390	1475	1028	1190
121	19,30,23,18,17,20,10	388,168,258,390,1475,1028,1190	20,29,21,15,14,19,12	1554,672,1032,1558,369,257,298	388	168	258	390	1475	1028	1190	1554	672	1032	1558	369	257	298
122	19,30,23,18,17,20,12	388,168,258,390,1475,1028,298	20,29,21,15,14,19,10	1554,672,1032,1558,369,257,1190	388	168	258	390	1475	1028	298	1554	672	1032	1558	369	257	1190
123	19,30,23,18,17,19,10	388,168,258,390,1475,257,1190	20,29,21,15,14,20,12	1554,672,1032,1558,369,1028,298	388	168	258	390	1475	257	1190	1554	672	1032	1558	369	1028	298
124	19,30,23,18,17,19,12	388,168,258,390,1475,257,298	20,29,21,15,14,20,10	1554,672,1032,1558,369,1028,1190	388	168	258	390	1475	257	298	1554	672	1032	1558	369	1028	1190
125	19,30,23,18,14,20,10	388,168,258,390,369,1028,1190	20,29,21,15,17,19,12	1554,672,1032,1558,1475,257,298	388	168	258	390	369	1028	1190	1554	672	1032	1558	1475	257	298
126	19,30,23,18,14,20,12	388,168,258,390,369,1028,298	20,29,21,15,17,19,10	1554,672,1032,1558,1475,257,1190	388	168	258	390	369	1028	298	1554	672	1032	1558	1475	257	1190
127	19,30,23,18,14,19,10	388,168,258,390,369,257,1190	20,29,21,15,17,20,12	1554,672,1032,1558,1475,1028,298	388	168	258	390	369	257	1190	1554	672	1032	1558	1475	1028	298
128	19,30,23,18,14,19,12	388,168,258,390,369,257,298	20,29,21,15,17,20,10	1554,672,1032,1558,1475,1028,1190	388	168	258	390	369	257	298	1554	672	1032	1558	1475	1028	1190

