



Universiteit
Leiden
The Netherlands

Current challenges in statistical DNA evidence evaluation

Cereda, G.

Citation

Cereda, G. (2017, January 12). *Current challenges in statistical DNA evidence evaluation*. Retrieved from <https://hdl.handle.net/1887/45172>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/45172>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/45172> holds various files of this Leiden University dissertation.

Author: Cereda, G.

Title: Current challenges in statistical DNA evidence evaluation

Issue Date: 2017-01-12

Current challenges in statistical DNA evidence evaluation

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op 12 januari 2017
klokke 15 uur

door

Giulia Cereda
geboren te Milano, Italië
in 1988

Promotores:

Prof. dr. R. D. Gill (Universiteit Leiden)

Prof. dr. F. Taroni (Université de Lausanne)

Samenstelling van de promotiecommissie:

Prof. dr. A. W. van der Vaart (Universiteit Leiden, chairman)

Prof. dr. P. Grünwald (Universiteit Leiden, secretary)

Prof. dr. A. Biedermann (Université de Lausanne)

Prof. dr. J. Mortera (Università di Roma tre)

Prof. dr. M. Sjerps (Universiteit van Amsterdam)

This thesis is part of a “cotutelle” agreement between the universities of Lausanne and Leiden. It has been supported by the Swiss National Science Foundation, through grants no. 105311-1445570 and 10531A-156146.

Alla mia famiglia

The front cover of this book was drawn by my father, Paolo Cereda.

In the back cover Lord Ganesha, the god of wisdom and learning, the patron of arts and sciences, as well as the remover of obstacles. One of his two tusks was broken to write down a very important textbook, the Mahabarata. According to some myths, he was generated by his mother, Parvathi, alone, using turmeric paste. For this reason, he represents women's independence.

A catalogue record is available from the Leiden University Library.

Preface

DNA profiling has become one of the most widely used techniques for human identification in forensic science since its introduction in 1984 by Alec Jeffreys. Despite the common belief that DNA evidence is a “damning evidence” which leaves no space for uncertainty, it actually needs strong statistical models in order to be used as a support for particular conjectures. The process which allows forensic experts to evaluate the statistical meaning of DNA evidence is one of the most interesting domains of forensic science of the last decades. This thesis started with the aim of building a statistical interpretative framework for a new genotyping methodology, the DIP-STR marker system, conceived to deal with the problem of extremely unbalanced mixtures.

While working on this project, we were confronted with the so-called ‘rare type match problem’, a very interesting open problem of forensic DNA statistics. The term refers to the situation in which there is a correspondence between the DNA profile of a suspect and that of a recovered stain, but this profile was never observed in a previously collected reference database. The evaluation of such a correspondence is very challenging. This problem is very common when using Y-STR markers or new genotyping techniques, such as DIP-STR markers, since the coverage of the available databases is limited. Therefore, we started investigating several statistical methods to deal with the rare type match problem. This led to the in-depth study of other delicate methodological issues, such as uncertainty assessment, data reduction, hybrid solutions.

As a closing loop to this Phd project, one of the discussed methods is proposed as a solution to the DIP-STR rare type match problem.

Contents

Preface	v
Introduction	1
Scope and Propositions	2
Novelty	3
Outline	3
I Background and summary	5
1 Preliminary concepts	7
1.1 Forensic DNA analysis	7
1.1.1 DNA as identification tool	7
1.1.2 Technical steps of DNA genotyping	10
1.1.3 Two classical techniques of DNA genotyping	11
1.1.4 DNA mixtures	12
1.1.5 Extremely unbalanced DNA mixture	13
1.2 Evaluation of DNA evidence	15
1.2.1 Bayesian inference and likelihood ratio	16
1.2.2 Frequentist likelihood ratio	17
1.2.3 Rare type match problem	17
1.2.4 Available solutions for the rare type match problem	18
1.2.5 The discrete Laplace method	19
1.3 Graphical models	20
1.3.1 Bayesian networks: formal definition	21
1.3.2 Object-orientation	23
1.3.3 Bayesian networks in forensic DNA literature	25
1.3.4 Object-oriented Bayesian networks for DNA evidence	26
1.4 Nonparametric Bayesian priors	29
1.4.1 The two-parameter Poisson Dirichlet distribution	29
1.4.2 Pitman sampling formula	30
1.4.3 The two-parameter Chinese restaurant process	30
1.4.4 The hyperparameters	31
1.4.5 Power law behavior	32
2 Results	33

2.1	Object-oriented Bayesian networks for evaluating DIP-STR profiling results from unbalanced DNA mixtures	33
2.2	An investigation of the potential of DIP-STR markers for DNA mixture analyses	35
2.3	Some methodological issues	37
2.4	Impact of model choice on LR assessment in case of rare haplotype match (frequentist approach)	40
2.5	A useful Lemma	42
2.6	Bayesian approach to LR for the rare match problem	43
2.7	Nonparametric Bayesian approach to LR assessment in case of rare haplotype match	45
2.8	A solution for the rare type match problem when using the DIP-STR marker system	47

II Papers 51

3	Object-oriented Bayesian networks for evaluating DIP-STR profiling results from unbalanced DNA mixtures	53
3.1	Introduction	54
3.2	Genetic background	55
3.3	An object-oriented Bayesian network (OOBN) for results of DIP-STR analyses	57
3.3.1	Evaluation of DNA profiling results using graphical models	57
3.3.2	The main class <code>Marker</code>	58
3.3.3	The main class <code>Marker for brother</code>	60
3.4	Casework examples	62
3.4.1	General case description and DIP-STR analyses	62
3.4.2	Case 1: suspect available	63
3.4.3	Case 2: missing suspect	68
3.4.4	A note on the likelihood ratio results	68
3.5	Discussion and conclusions	69
3.6	Acknowledgements	70
4	An investigation of the potential of DIP-STR markers for DNA mixture analyses	75
4.1	Introduction	76
4.2	Genetic background	77
4.2.1	DIP-STR markers	77
4.2.2	STR markers	78
4.2.3	Y-STR markers	79
4.3	Interpretative model	79
4.3.1	Likelihood ratios for STR markers	81
4.3.2	Likelihood ratios for DIP-STR markers	82
4.3.3	Likelihood ratios for the Y-STR markers	83
4.4	Comparison of the three methods	84
4.4.1	Comparison of DIP-STR and STR assuming point of view of the prosecution	85

4.4.2	Comparison between DIP-STR and STR marker systems assuming the point of view of the defence	87
4.4.3	Comparison between DIP-STR and Y-STR marker systems assuming the point of view of the prosecution	88
4.4.4	Comparison between DIP-STR and Y-STR marker systems assuming the point of view of the defence	89
4.4.5	A discussion about the influence of genetic model assumptions	90
4.5	Consideration on the usefulness of the three methods	90
4.6	Conclusion	92
5	Impact of model choice on LR assessment in case of rare haplotype match (frequentist approach)	97
5.1	Introduction	97
5.2	Bayesian versus frequentist approach to likelihood ratio assessment	99
5.2.1	The Bayesian approach	100
5.2.2	The frequentist perspective	101
5.3	Data reduction	102
5.4	Different levels of uncertainty	103
5.4.1	Estimating the weight of evidence	104
5.5	The rare Y-STR haplotype problem	105
5.6	The Discrete Laplace Method	106
5.6.1	The choice of D in the Discrete Laplace Method	107
5.6.2	Quantifying the uncertainty of the Discrete Laplace method	108
5.7	The Generalized Good method	110
5.7.1	Quantifying the uncertainty of the GG method	112
5.8	Choosing and comparing methods	114
5.9	Remark and conclusion	114
6	Bayesian approach to LR for the rare match problem	117
6.1	Introduction	117
6.1.1	Notation	119
6.2	Genetic terminology	119
6.3	The rare type match problem	120
6.4	The full Bayesian approach to LR	121
6.4.1	Bayesian point of view.	122
6.4.2	Frequentist point of view.	123
6.4.3	The Bayesian plug-in LR and the proper Bayesian LR	124
6.4.4	State of the art for DNA match evaluation	124
6.5	A useful Lemma	126
6.6	Bayesian LR calculation, based on beta-binomial model	127
6.7	Bayesian LR calculation, based on Dirichlet-multinomial model	129
6.7.1	Poisson prior	131
6.7.2	Negative binomial prior	133
6.7.3	Sensitivity analysis	135
6.7.4	Remarks about conventional priors	135
6.8	Conclusion	136

7	Nonparametric Bayesian approach to LR assessment in case of rare type match	139
7.1	Introduction	139
7.2	A Bayesian nonparametric model for the rare type match	140
7.2.1	The rare type match problem	140
7.2.2	Notation	142
7.2.3	Model assumptions	142
7.2.4	Prior	143
7.3	The model	144
7.3.1	Random partitions	146
7.3.2	Chinese Restaurant representation	148
7.4	Some results	149
7.4.1	A useful Lemma	149
7.4.2	Known results about the two-parameter Poisson Dirichlet distribution	151
7.5	The likelihood ratio	152
7.5.1	True LR	152
7.6	Analysis on a real database	155
7.6.1	Model fitting	155
7.6.2	Loglikelihood	156
7.6.3	Analyzing the error	157
7.7	Conclusion	159
8	A solution for the rare type match problem when using the DIP-STR marker system	161
8.1	Introduction	162
8.2	DIP-STR marker system for extremely unbalanced mixtures	163
8.3	Bayesian network for evaluating DIP-STR profiling results from unbalanced DNA mixtures.	164
8.4	Notation	165
8.5	Full Bayesian approach	166
8.6	Rare type match problem	167
8.7	A prior for θ	167
8.8	Full model	169
8.9	Lemma	170
8.10	Choice of priors	172
8.10.1	Alternative solutions	172
8.11	Conclusion	174
III	Discussion	181
9	Discussion and conclusion	183
9.1	Contribution to the practice of Forensic Science	183
9.2	Contribution to the Philosophical point of view	184
9.3	Future perspective	184
9.4	Conclusion	185