

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/130241>

Please be advised that this information was generated on 2019-10-21 and may be subject to change.

The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

Jacob de Zoete · Marjan Sjerps · Ronald Meester ·
Eric Cator.

Received: date / Accepted: date

Abstract When a Y-chromosomal and a (partial) autosomal DNA profile are obtained from one crime sample, and both profiles match the suspect's profiles, we would like to know the combined evidential value. To calculate the likelihood ratio of observing the autosomal and Y-chromosomal DNA profiles combined, we need to know the conditional random match probability of the observed autosomal DNA profile, given the Y-chromosomal match. We examine this conditional probability in two ways: (1) with a database containing 2085 men, and (2) using a simulation model. We conclude that if the Y-chromosomal DNA profiles match, we can still regard the autosomal DNA profile as independent from the Y-chromosomal DNA profile if the matching person is not a descendant of the father of the donor of the (crime) sample. The evidential value can in that case be computed by multiplying the random match probabilities of the individual profiles.

Keywords random match probability · Y-STR · combined evidential value · forensic DNA statistics · dependence Y and autosomal loci

1 Introduction

Suppose that a crime stain is found at a crime scene, and that two DNA profiles are obtained from it: a Y-chromosomal DNA profile and a partial autosomal DNA profile. Furthermore, no indication of a mixture is observed so we assume that the two profiles originate from the same person. When a suspect is identified that matches both profiles, the question arises how to assess the combined evidential value. Following the likelihood ratio approach described in e.g. Evett and Weir [1], we define the hypotheses H_p : *the suspect is the donor of the crime stain*, and H_d : *some other man is the donor of the crime stain*. Usually, this other man is assumed to be unrelated to the suspect. We will not make this assumption here. The evidential value of the match is then expressed as the likelihood ratio (LR) of the observations

Jacob de Zoete
University of Amsterdam, Korteweg de Vries Instituut voor Wiskunde
Postbus 94248, 1098 XH Amsterdam, The Netherlands
E-mail: j.c.dezoete@uva.nl

Marjan Sjerps
Netherlands Forensic Institute
Laan van Ypenburg 6 2497 GB Den Haag, The Netherlands
Postbus 24044, 2490 AA Den Haag, The Netherlands

Ronald Meester
VU University Amsterdam, Department of Mathematics
De Boelelaan 1081a, 1081 HV, Amsterdam, The Netherlands

Eric Cator
Radboud University Nijmegen, Institute of Mathematics, Astrophysics and Particle Physics
Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

(the autosomal and Y-chromosomal DNA profiles of the stain and of the suspect), considering these hypotheses. It is easy to derive that this LR simplifies to $1/(p \cdot q)$, where p equals the probability that the observed Y-chromosomal profile of the stain matches that of the suspect, and q equals the conditional probability that the observed autosomal DNA profile of the stain matches that of the suspect, *given the Y-chromosomal match*.

Combining the evidential value of the DNA profiles by simply multiplying the Y-chromosomal and autosomal random match probabilities can only be done when the profiles can be assumed to be independent, which is not a priori obvious. Indeed, the observation that two men share the same Y-chromosome makes it more likely that they are related, and this makes it more likely that their autosomal profiles will match.

If the population of alternative suspects contains no (or only distant) relatives of the suspect, we can assume independence (conditional on H_d) between the autosomal and Y-chromosomal DNA profiles and q simplifies to the “standard” random autosomal match probability. If the population of alternative suspects contains relatives of the suspect but we know the number of relatives as well as their degree of relatedness, we can also compute the likelihood ratio (by using the Weight-of-Evidence Formula from [2], see also Anderson and Weir [3] and Bright, Curran and Buckleton [4]). Unfortunately, in practice we usually do not know the number of relatives, nor their order of relatedness, and in this paper we will investigate this situation.

There are several papers in the literature which discuss the combination of autosomal and Y-chromosomal DNA profiles. Amorim [5] states that combining the evidential value obtained from lineage markers (such as Y-chromosome) with that resulting from individuality markers (autosomal) is difficult, if not impossible. Amorim advises to change the prosecutor’s hypothesis from *The suspect is the source of the stain* to *The suspect or somebody from their lineage is the source of the stain* and to report the analysis from the two types of evidential sources separately, along with the distinct theoretical and statistical frameworks underlying them. In that way, it will always be possible to join the two types of evidence, but the assumption of non-involvement of relatives will have to be explicitly accepted by the court in the case under judgment. However, in [6], Weir et al. state that there seems to be neither a logical nor a legal basis for changing the prosecution hypothesis like Amorim suggests. Moreover, they remark that if the possibility of mutation is neglected, then the likelihoods of observing the profiles given the two hypotheses would be identical for lineage markers anyway.

Walsh, Redd and Hammer [7] give several complications in computing Y-autosomal joint match probabilities. Apart from the fact that two individuals sharing the same Y haplotype are likely to be more closely related than two random individuals from the populations, they remark that Y chromosome haplotypes may be highly informative as to which sub-population an individual belongs, and this in turn may change the autosomal allele frequencies used to compute the autosomal match probabilities. They performed an independence test on autosomal and Y STRs on 16 populations. Their results did not give any reason to reject independence of Y and autosomal markers. They recommend to compute the joint Y-autosomal matching probability by computing the autosomal match probability as the product of single-locus genotype frequencies. These are corrected by using the sampling formula of Balding and Nichols, [8], with $\theta = 0.04$. This can be done for several populations, and finally the maximum of the resulting random match probabilities is multiplied with the estimated matching probability for the Y haplotype to obtain the joint random match probability. This recommendation is a conservative approach to computing the joint Y-autosomal matching probability. It is based on Walsh et al.’s calculations showing that when population structure is already present in the autosomes, the additional effect due to conditioning on the Y is small.

Budowle et al. [9] also performed an independence test on autosomal and Y STRs, for three sampled populations of unrelated males from Texas. The test did not give any reason to reject that the frequencies of autosomal and Y STR profiles can be combined using the product rule.

None of these papers, however, solves our problem. In most cases, the method is not applicable for a population containing many relatives. In some of the papers, a database is used to investigate dependence between autosomal and Y-chromosomal DNA profiles. The databases used are quite small, and when the conclusion of the database research is that there is no reason to assume dependence we cannot be sure whether there actually exists dependence in the population. The power of the statistical test is very important here, see Validating Databases in [10]. The suggestion of using $\theta = 0.04$ on the population with the maximum Y match probability has similarities with the ceiling principle, and may be a source of criticism [11].

In this paper we examine the dependence between matching autosomal and matching Y-chromosomal DNA profiles in two different ways. First we investigate dependence in a data set of 2085 Dutch males (blood donors who volunteered for forensic research, de Knijff and Sijen, in preparation) gathered by The Forensic Laboratory for DNA Research of the Leiden University Medical Center. The data shows significant departures from independence. However, this can be due to several causes which we explore. Secondly, we investigate the impact of the dependence with a general simulation model which gives us the opportunity to assess the consequences for forensic case work. It turns out that for relatives that do not have the father of the suspect as an ancestor, the dependence is negligible, and we can safely compute the random match probability for the two combined profiles by multiplying the random match probabilities of the individual profiles. A separate issue is how to compute the random match probabilities of Y-chromosomal DNA profiles in the first place. We will briefly discuss this problem in Section 5.

2 The database independence test - method and result

In this section we investigate the dependence between autosomal and Y-chromosomal markers in a database containing 2085 men. Our database has information on 23 autosomal markers and 17 Y-chromosomal loci. With 2085 men, we can compare $\binom{2085}{2} = 2,172,570$ pairs of men. For each of these pairs of men and for each autosomal marker, we consider the number of autosomal alleles that match (0, 1 or 2). Furthermore, we consider whether or not their Y-chromosomal DNA profiles match (on 17 loci). Thus, we can make 23 contingency tables of 2 by 3. An example of such a table is given in Table 1.

Table 1 Contingency table for autosomal locus D1S1656

	autosomal alleles			Total
	0 matching	1 matching	2 matching	
no matching Y-haplotype	1447795	685267	39067	2172129
matching Y-haplotype	281	149	11	441
Total	1448076	685416	39078	2172570

We calculated the chi-square test statistic for each contingency table. Since we are comparing all possible pairs of men, the observations in the contingency table are not independent (a match between man 1 and 2 and a match between man 1 and 3 means there will be a match between man 1 and 3). Therefore, similar to [7,12,13], the distribution of the chi-square test statistic, assuming independence, was determined by a permutation test. We constructed new databases containing 2085 men by making permutations of the auto and Y-chromosomal DNA profiles in our database.

This way, we found 5 autosomal loci (D2S1338, SE33, TH01, D12S391, D13S317) with approximated p -values less or equal to 0.05 of which 1 was below 0.01 (SE33). The probability of finding at least 5 p -values less or equal 0.05 in 23 experiments assuming that the observations are independent is approximately 1 in 200, hence, we have reason to believe that there may be dependence between the markers. However, there are several explanations for the low p -values: (1) there actually exists dependence between autosomal and Y-chromosomal markers, or (2) the database contains relatives. Due to the way the database was constructed (using blood donors), it is not unlikely that the latter is the case. Especially on loci with high allelic variation (SE33), relatives in the database can have major influence on observing dependence between autosomal and Y-chromosomal DNA profiles. Indeed, The Forensic Laboratory for DNA Research of Leiden University Medical Center confirmed that there are at least three father-son pairs. It is not unlikely that there are more relatives in the database.

In Figure 1 we give boxplots with the number of matching autosomal alleles for all the pairs of men given the number of matching Y-chromosomal loci. We see that the boxplots are similar for 0-14 matching Y-chromosomal loci. When the number of matching Y-chromosomal loci is larger than 14 we start to find large outliers. The red line represents the mean number of matching autosomal alleles as a function of the number of matching Y-chromosomal loci. We see that it is approximately constant except for completely matching Y-chromosomal DNA profiles (17 loci). Again, it is possible that the large outliers represent pairs of closely related men, where in some cases there is a mutation on the Y-chromosome, but it is also possible that something else is causing the outliers.

127 The relevant forensic question is: what is the combined evidential value of matching Y-chromosomal
128 and autosomal DNA profiles? Simple multiplication is not straightforward, because two men with match-
129 ing Y-chromosomal DNA profiles may be more likely to have matching autosomal DNA profiles. The
130 database provides some evidence to support the preceding statement but it is not very informative. The
131 number of autosomal matches in a population of people that may be related to the suspect is something
132 that can be easily investigated with a simulation model.

133 3 The simulation - method

134 We are interested in the combined evidential value of matching Y-chromosomal and autosomal DNA
135 profiles when the population of alternative suspects contains an unknown number of relatives of the
136 suspect. A conservative assumption (defined as an assumption that is in favor of the suspect) is that all
137 men in this population are related to the suspect. If we assume no mutations (which is again a conservative
138 assumption) then all these men share the Y-chromosome of the suspect. The autosomal match probability
139 for an individual in this population is an upper bound for the autosomal match probability in more realistic
140 alternative populations. Hence, our goal is to derive the autosomal match probability in a population
141 consisting of relatives of the suspect. We will compare this with the standard autosomal random match
142 probability in an unrelated population. If the difference is negligible we can compute the likelihood ratio
143 of the two profiles combined by multiplying the likelihood ratios of the individual profiles.

144 We construct such populations by simulating family trees that relate a given number n of men to the
145 suspect. We assume that all the men in this tree share the same Y-chromosome. To simulate family trees,
146 we assume that the number of sons each man fathers is binomially distributed with mean 1.3. This is
147 slightly conservative in view of the world average 1.24 [14]. We take the last person that is added to the
148 tree, the n th living person, as our suspect. This has no effect on our conclusion because we will focus on
149 members of the population that are not descendants of the father of the suspect.

150 By assigning the suspect a randomly drawn (partial) autosomal DNA profile and by assuming Mendelian
151 inheritance, we can assign (partial) autosomal DNA profiles to all the men in the family tree. The (par-
152 tial) autosomal DNA profiles of the mothers in the tree are randomly drawn. The allele frequencies used
153 are computed using the same database, containing 2085 men, as in Section 2. A detailed description of
154 the simulation model can be found in the supplementary material. This can be found in the online version
155 of this paper.

156 We simulated the autosomal DNA profile assignment multiple times over the same tree, always starting
157 with the same (partial) autosomal DNA profile for the suspect, to estimate the distribution of the number
158 of matching profiles as well as the match probability in the family tree. This is done with a large number of
159 different family trees. We compared the distributions of the number of matching profiles of each individual
160 tree with each other. Furthermore, to investigate the influence of relatedness on the number of matches,
161 we compare the distributions for the family trees with the distribution of an unrelated population (which
162 is a binomial distribution).

163 4 The simulation - results

164 We carried out a large number of simulations. We present a restricted number of results here; more results
165 can be found in the supplementary material. The parameter settings of the situations that are presented
166 here are given in Table 2. For each simulation, at least 100 different family trees were simulated. This
167 way, we can examine the influence of the shape of the family tree on the results. We are interested in the
168 differences between the distribution of the number of matching profiles in a related population (family
169 tree) and in an unrelated population (binomial distribution).

170 Generally, we see that they differ most in the tails of the distributions: It is more likely to see a larger
171 number of matching profiles as well as smaller number of matching profiles in related populations than
172 in unrelated populations. This is to be expected since in a family tree, there is correlation between the
173 autosomal DNA profiles.

174 For each situation in Table 2, we selected a tree whose distribution deviated most in the mean and tails
175 with the binomial distribution corresponding to an unrelated population. The tree with the distribution
176 that deviated most from the distribution for an unrelated population in the first situation given in Table

Table 2 Overview of the parameters we used in the simulations

number of men	$\mathbb{E}[\text{sons}]$	number of autosomal loci	random match probability	Figure
1000	1.29	3	$1.75 \cdot 10^{-3}$	2
100	1.29	2	$2.69 \cdot 10^{-2}$	3
10	1.29	2	$2.69 \cdot 10^{-2}$	4
1000	2.58	3	$1.75 \cdot 10^{-3}$	5
1000	1.29	6	$5.18 \cdot 10^{-7}$	6
1000	1.29	1	$6.58 \cdot 10^{-4}$	7

2 is given in blue in Figure 2. With these parameters, the expected number of matching profiles in an unrelated population would be 1.75. In this tree, the expected number of matching profiles is 2.35. It turns out that this difference is mainly due to the close relatives of the suspect. If we only consider the men that do not have the father of the suspect as an ancestor (so we do not omit the profiles of the father and the brothers of the suspect) then we get the distribution given in green in Figure 2. The mean number of matching profiles in this family tree, where one branch is “cut off”, is 1.77. This makes sense, and is typical for all the trees we simulated.

This suggests that it may be an option to change the alternative hypothesis that we consider from *some other man is the donor of the crime stain to some other man, who is not a descendant of the father of the suspect, is the donor of the crime stain*. This means that the focus shifts to the family tree with this branch cut off.

To compare the distributions, we compare the mean and the 95th percentile of the distribution of the trees and the binomial distribution, see Table 3. The mean and 95th percentile of the related population distributions grow compared to 95th percentile of unrelated populations when we decrease number of men in the alternative population. A smaller number of men in the alternative population is similar to increasing the proportion of men that is closely related to the suspect. A tree relating a billion men has a negligible number of men who are closely related to the suspect. The majority of men in such a tree have autosomal DNA profiles that can be regarded as independent of the autosomal DNA profile of the suspect, since their common ancestor is very distant.

Table 3 Mean and 95th percentile in related and unrelated populations for the different simulations where the father branch is cut from the family tree

number of men	$\mathbb{E}[\text{sons}]$	number of loci	random match probability	unrelated		related; all family trees		related; maximum over all trees	
				mean	95th percentile	mean	95th percentile	mean	95th percentile
1000	1.29	3	$1.75 \cdot 10^{-3}$	1.75	4	1.78	5	1.91	5
100	1.29	2	$2.69 \cdot 10^{-2}$	2.69	6	2.79	7	3.05	8
10	1.29	2	$2.69 \cdot 10^{-2}$	0.27	1	0.35	2	0.44	3
1000	2.58	3	$1.75 \cdot 10^{-3}$	1.75	4	1.85	5	1.97	6
1000	1.29	6	$5.18 \cdot 10^{-7}$	0.08	1	0.09	1	0.10	1
1000	1.29	1	$6.58 \cdot 10^{-4}$	0.66	2	0.51	2	0.63	3

Increasing the expected number of sons a man gets has a similar effect. The proportion of closely related men increases when the expected number of sons increases. The number of generations that is needed to relate a fixed number of men decreases. However, the effect of increasing the expected number of sons or decreasing the number of men in the alternative population is small. The mean number of matching autosomal as well and the 95th percentile of the distributions do increase, but the difference is small, see Table 3. It is important to note that both the situation where the alternative suspect population consists of 10 men and the situation where the expected number of sons a man gets is 3.87 are unrealistic. In the simulations that represent realistic situations (i.e. a normal expected number of

sons, realistic number of men in the alternative population) the difference between the mean number of matching autosomal DNA profiles in an unrelated and in a related population (without the men in the father branch of the tree) are very small. As said, the differences in distribution occur in the tails of the distribution, as can be seen in the 95th percentiles, but these 95th percentiles do not substantially differ.

More results and a sensitivity analysis on the different parameters can be found in the supplementary material, which can be found at in the online version of this paper.

5 Conclusion and Discussion

A test for independence for Y-chromosomal DNA profiles and autosomal loci on the Dutch database containing 2085 men yielded a significant result on 5 of the 23 autosomal loci. However, there are multiple possible explanations for this. For instance, it can be due to close relatives in the database.

Our simulation model showed that although the distribution of the number of matching profiles for an unrelated and a related population are not the same, the expected numbers of matching profiles are very close to each other. Especially when we exclude some of the close family of the suspect (everybody that has the father of the suspect as an ancestor), our results indicate that the difference between the expected number of matching profiles in a related and an unrelated population is very small. The differences in the 95th percentiles of the distributions are more pronounced, but they remain small. The results show that we can regard autosomal and Y-chromosomal DNA profiles as independent when we exclude the relatives of the suspect that have the father of the suspect as an ancestor.

We suggest to compute the likelihood ratio, in a case where an autosomal and a Y-chromosomal DNA profile are obtained from the same stain, using the hypotheses H_p : *the suspect is the donor of the crime stain* and H_d : *some other man, who is not a descendant of the father of the suspect, is the donor of the crime stain*. When doing so, the likelihood ratio can be computed by multiplying the individual likelihood ratios of the autosomal and Y-chromosomal DNA profiles.

As mentioned, in [7], Walsh et al. suggest to use $\theta = 0.04$ in the sampling formula of Balding and Nichols [8] to estimate the random match probability of the autosomal DNA profile. This value is multiplied with the estimated matching probability of the Y-chromosomal DNA profile to obtain the joint match probability. This approach will give larger joint random match probabilities than our approach. This is because Walsh et al. consider the alternative donor as a random draw from a homogeneous population, that may include close relatives. We provide a match probability for the population not including the close relatives of the suspect. The match probability of close relatives can be calculated and reported as well [1,3,10]). We think our approach describes the situation more accurately because the alternative donor population is not homogeneous. Close relatives of the suspect have a significantly higher match probability. It is important the the court is informed about this.

We are aware of the scientific discussion on how to compute the random match probability of a Y-chromosomal DNA profile [15,16,17,18,19]. Whatever the outcome of this discussion may be, the results of this paper can still be used to combine the evidential value of the two profiles.

This paper focuses on the combined evidential value of the two DNA profiles. However, the conclusions from this paper are also relevant in a wider context; in reporting autosomal matches. Currently, these are reported with an alternative hypothesis stating that the crime stain was left by an unknown unrelated person. In a situation where a crime was committed in a small village, it is likely that the alternative suspect population contains many relatives of the suspect. In such a case, the conclusion that it is a million times more likely to find the DNA match when the suspect is the donor of the crime stain than when some unrelated unknown person was the donor has little influence. The judge has no idea what the actual likelihood ratio is since the alternative suspect population consists of relatives of the suspect. Now, we can use the conclusions from our simulation model and suggest a new alternative hypothesis H_d *some other person, who is not a descendant of the father of the suspect, is the donor of the crime stain* and we can compute the likelihood ratio by multiplying the allele frequencies of the alleles in the DNA profile of the crime stain. However, we need to know the gender of the donor to do so.

References

1. Ian W. Evett and Bruce S. Weir. *Interpreting DNA evidence: statistical genetics for forensic scientists*. Sinauer Associates, 1998.

2. David J. Balding. *Weight-of-Evidence for Forensic DNA Profiles*. John Wiley and Sons, 1995. 255
3. Amy D. Anderson and Bruce S. Weir. It was one of my brothers. *International Journal of Legal Medicine*, 120(2):95–104, March 2006. 256
4. Jo-Anne Bright, James Michael Curran, and John S. Buckleton. Relatedness calculations for linked loci incorporating subpopulation effects. *Forensic Science International. Genetics*, 7(3):380–3, May 2013. 257
5. António Amorim. A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. *Forensic Science International. Genetics*, 2(4):376–8, September 2008. 258
6. John S. Buckleton, Michael Krawczak, and Bruce S. Weir. The interpretation of lineage markers in forensic DNA testing. *Forensic Science International. Genetics*, 5(2):78–83, March 2011. 259
7. Bruce Walsh, Alan J Redd, and Michael F Hammer. Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Science International*, 174(2-3):234–8, January 2008. 260
8. David J Balding and A Nichols. DNA profile match probability calculation: how to allow for population stratification , relatedness , database selection and single bands. *Forensic Science International*, 64:125–140, 1994. 261
9. Bruce Budowle, Jianye Ge, Xavier G. Aranda, John V. Planz, Arthur J. Eisenberg, and Ranajit Chakraborty. Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence*. *Journal of Forensic Sciences*, 54(5):1016–21, September 2009. 262
10. John S. Buckleton, Simon J. Walsh, and Christopher M. Triggs. *Forensic DNA Evidence Interpretation*. CRC Press, 2005. 263
11. David H. Kaye. *The Double Helix and the Law of Evidence*. Harvard University Press, 2010. 264
12. Bruce Budowle, Mike Adamowicz, Xavier G. Aranda, Charles Barna, Ranajit Chakraborty, Dan Cheswick, Bradley Dafoe, Arthur J. Eisenberg, Roger Frappier, Ann Marie, Carll Ladd, Hee-suk Lee, Scott C Milne, Carole Meyers, Mechthild Prinz, Melanie L Richard, Gabriela Saldanha, Amy A Tierney, Lori Viculis, and Benjamin E Krenke. Twelve short tandem repeat loci Y chromosome haplotypes : Genetic analysis on populations residing in North America. *Forensic Science International*, 150:1–15, 2005. 265
13. Sudhir K Sinha, Bruce Budowle, Ranajit Chakraborty, Ana Paunovic, Robin Deville Guidry, Chris Larsen, Amrita Lal, Megan Shaffer, Gina Pineda, Siddhartha K Sinha, Elaine Schneida, and Huma Nasir. Utility of the Y-STR Typing Systems Y-PLEX TM 6 and Y-PLEX TM 5 in Forensic Casework and 11 Y-STR Haplotype Database for Three Major Population Groups in the United States . *Journal of Forensic Sciences*, 49(4):1–10, 2004. 266
14. Central Intelligence Agency. *The World Factbook*. Directorate of Intelligence, 2012. 267
15. Charles H Brenner. Fundamental problem of forensic mathematics—the evidential value of a rare haplotype. *Forensic Science International. Genetics*, 4(5):281–91, October 2010. 268
16. Thore Egeland and Antonio Salas. Estimating haplotype frequency and coverage of databases. *PloS one*, 3(12):e3988, January 2008. 269
17. Mikkel Meyer Andersen. *Haplotype Frequency Estimation and Master of Science Thesis*. PhD thesis, Aalborg University, 2010. 270
18. Mikkel Meyer Andersen, Amke Caliebe, Arne Jochens, Sascha Willuweit, and Michael Krawczak. Estimating trace-suspect match probabilities for singleton Y-STR haplotypes using coalescent theory. *Forensic Science International. Genetics*, 7(2):264–71, February 2013. 271
19. L Roewer, M Kayser, P de Knijff, K Anslinger, A Betz, A Caglià, D Corach, S Füredi, L Henke, M Hidding, H J Kärger, R Lessig, M Nagy, V L Pascali, W Parson, B Rolf, C Schmitt, R Szibor, J Teifel-Greding, and Michael Krawczak. A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Science International*, 114(1):31–43, October 2000. 272

Acknowledgements We wish to thank Peter de Knijff from the Leiden University Medical Center, for providing us with the database consisting of 2085 men to investigate the dependence between autosomal and Y-chromosomal DNA profiles. This work is part of the project “Combining Evidence” in the Forensic Science programme [727.011.007], financed by the Netherlands Organisation for Scientific Research (NWO). This study was supported by a grant from the NWO within the framework of the Forensic Genomics Consortium Netherlands. 297

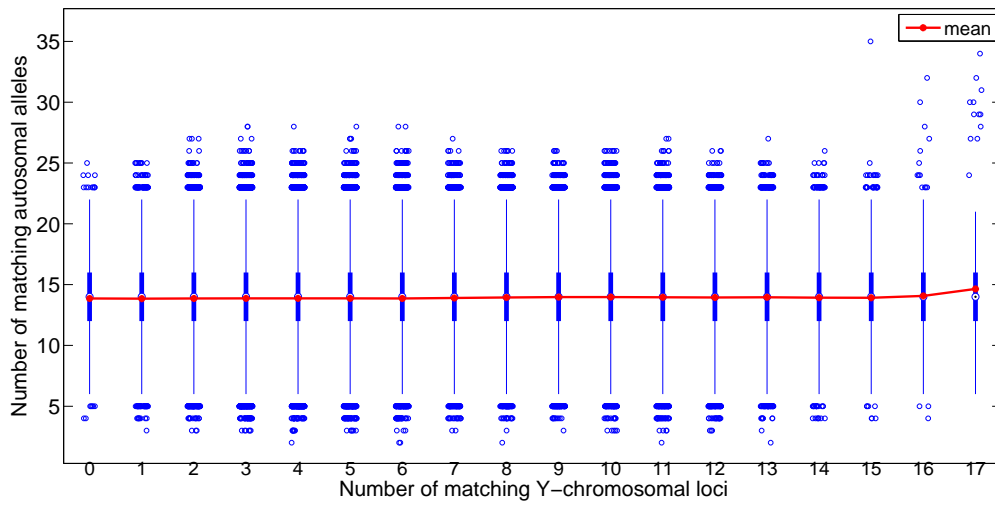


Fig. 1 Boxplots for the number of autosomal alleles that match between pairs of men, given the number of matching Y-chromosomal loci

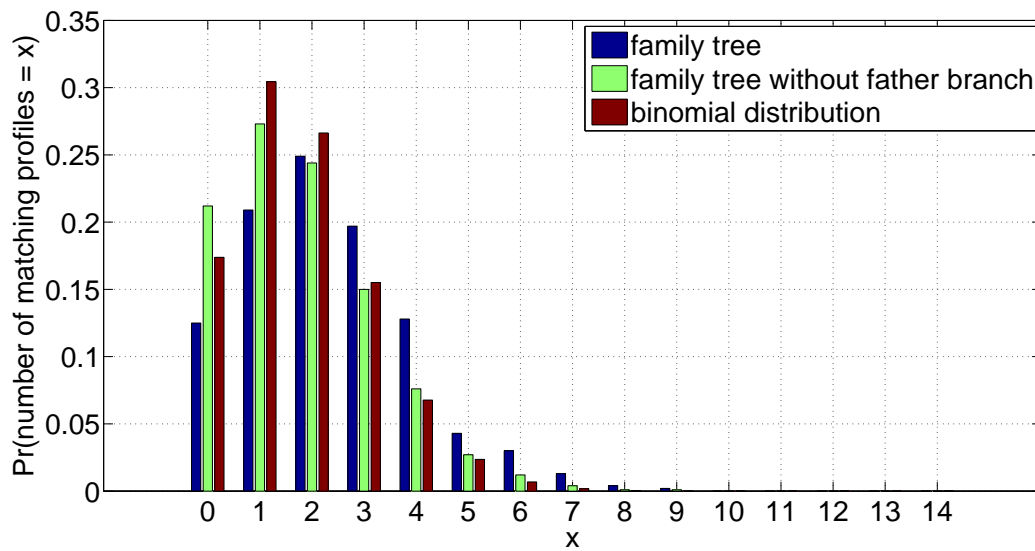


Fig. 2 pdf of the number of matching autosomal DNA profiles in a tree relating 1000 men, random match probability 0.0017, $\mathbb{E}[\text{sons}] = 1.29$ and a partial autosomal DNA profile on 3 loci as the suspect profile

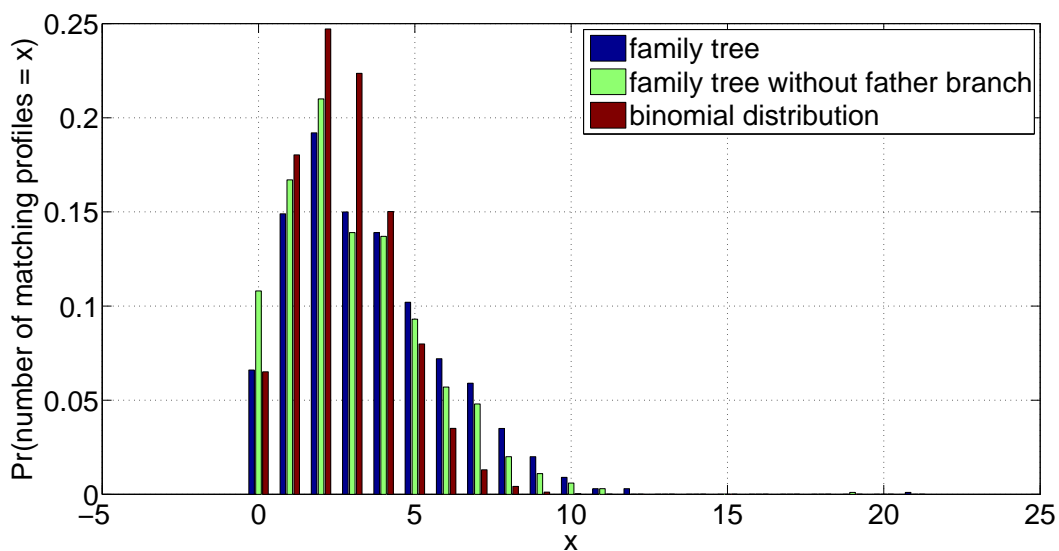


Fig. 3 pdf of the number of matching autosomal DNA profiles in a tree relating 100 men, random match probability 0.0269, $\mathbb{E}[\text{sons}] = 1.29$ and a partial autosomal DNA profile on 2 loci as the suspect profile

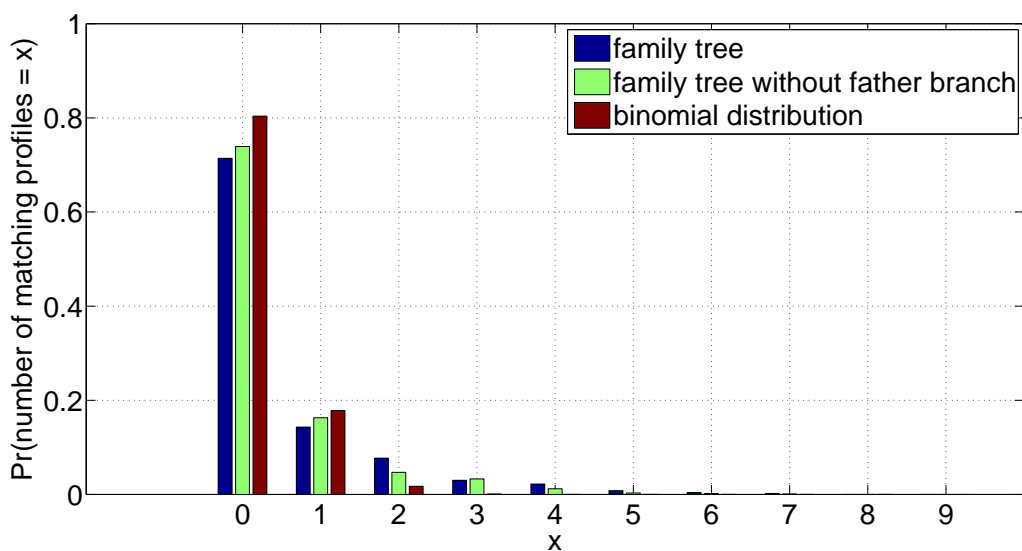


Fig. 4 pdf of the number of matching autosomal DNA profiles in a tree relating 10 men, random match probability 0.0269, $\mathbb{E}[\text{sons}] = 1.29$ and a partial autosomal DNA profile on 2 loci as the suspect profile

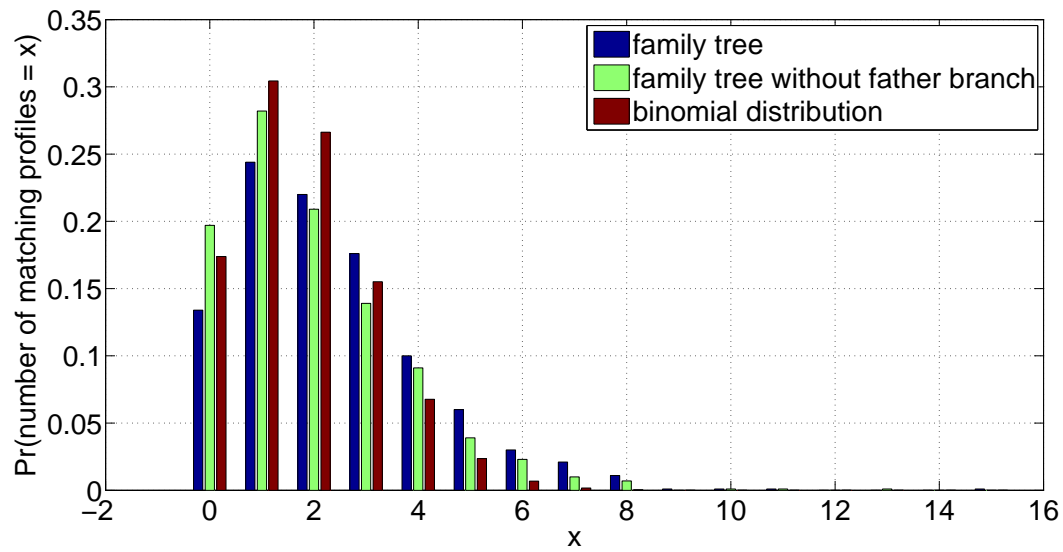


Fig. 5 pdf of the number of matching autosomal DNA profiles in a tree relating 1000 men, random match probability 0.0017, $\mathbb{E}[\text{sons}] = 2.58$ and a partial autosomal DNA profile on 6 loci as the suspect profile

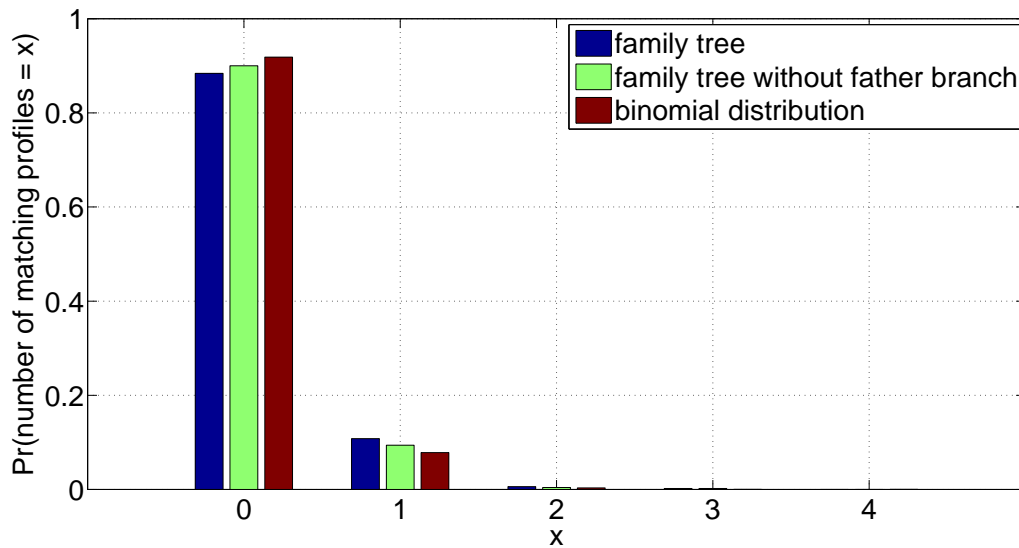


Fig. 6 pdf of the number of matching autosomal DNA profiles in a tree relating 1000 men, random match probability $8.52 \cdot 10^{-5}$, $\mathbb{E}[\text{sons}] = 1.29$ and a partial autosomal DNA profile on 6 loci as the suspect profile

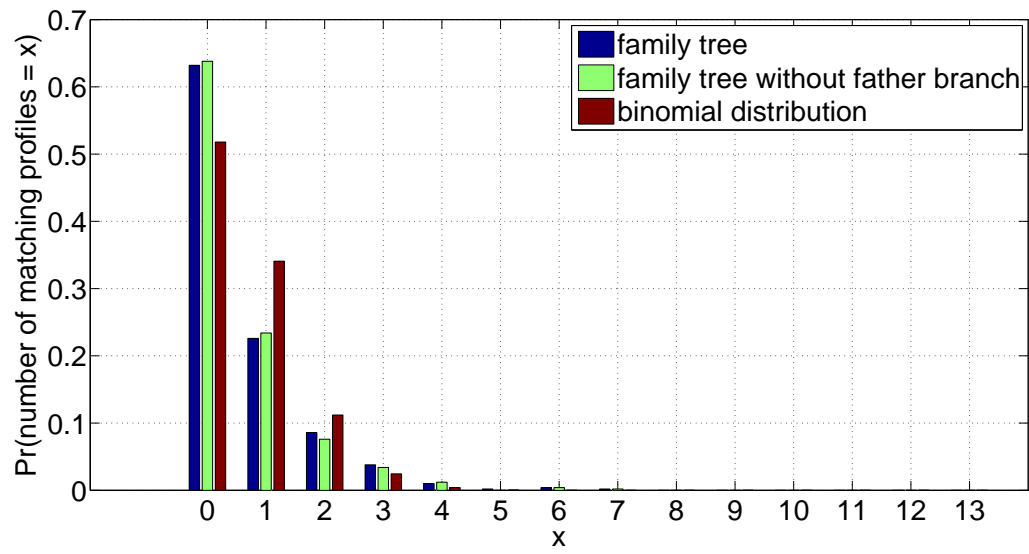


Fig. 7 pdf of the number of matching autosomal DNA profiles in a tree relating 1000 men, random match probability $6.58 \cdot 10^{-4}$, $E[\text{sons}] = 1.29$ and a partial autosomal DNA profile on 1 locus as the suspect profile