How convincing is a matching Y-chromosome profile?

Mikkel M Andersen[1,2] and David J Balding[3,4,*]

[1]Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark
[2]Section of Forensic Genetics, Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark
[3]Centre for Systems Genomics, Royal Parade, University of Melbourne, Vic 3010, Australia
[4]Genetics Institute, University College London, Gower St, WC1E 6BT, UK
[*]Corresponding author: dbalding@unimelb.edu.au

April 28, 2017

Abstract

The introduction of forensic autosomal DNA profiles was controversial, but the problems were successfully addressed, and DNA profiling has gone on to revolutionise forensic science. Y-chromosome profiles are valuable when there is a mixture of male-source and female-source DNA, and interest centres on the identity of the male source(s) of the DNA. The problem of evaluating evidential weight is even more challenging for Y profiles than for autosomal profiles. Numerous approaches have been proposed; we explain why they are unsatisfactory. Men with matching Y-profiles are related in extended patrilineal clans, many of which may not be represented in available databases. The relevant population is difficult to define, and the number of matching relatives is fixed as population size varies, making it infeasible to derive population-based match probabilities relevant to a specific crime. We propose a surprisingly simple solution to the problem. We provide a simulation model and software to approximate the distribution of the number of males with matching Y profiles, and use importance sampling reweighting to condition on a count from a randomly-sampled database if available. Our approach allows, for the first time, the value of Y profile evidence to be quantified for a court in a way that is both scientifically valid and easily comprehensible by a judge or juror.

Contents

1 Introduction **2**

2 Methods **4**
2.1 Marker sets and mutation 4
2.2 Population and mutation simulations 6
2.3 Comparison of simulated and real databases 6
2.4 Identifying matching males 7
2.5 Standard deviation in number of male offspring 7
2.6 Importance weighting to condition on database match count 8

3 Results **9**
3.1 Comparison of simulated and real databases 9
3.2 The number of matching Y profiles 10
3.3 Effect of standard deviation of offspring number 10
3.4 Role of database information 11

4 Discussion **12**
4.1 Presentation in court 12
4.2 Coancestry, kinship and sampling adjustments 13
4.3 Likelihood ratio reporting 13
4.4 Mitochondrial DNA profiles 14
4.5 Software 14
4.6 Acknowledgments 14

A Supplementary Text: Mixtures **16**

B Supplementary figures **17**

C Supplementary tables **21**

1 Introduction

A forensic Y-chromosome profile typically consists of the allele at between 15 and 30 short tandem repeat (STR) loci. For autosomal STR profiles, there are two alleles per locus and because of the effects of recombination, the alleles at distinct loci are treated as independent, after any adjustments for sample size, coancestry and direct relatedness. Y profiles usually have only one allele per locus, the exceptions are due to genomic duplications.[1] The loci lie in the non-recombining part of the Y chromosome, which is in effect a single locus and so the Y profile can be regarded as a single allele, or haplotype. This represents a major contrast with autosomal profiles, the implications of which have not been fully appreciated.[2]

At the core of the weight of evidence for autosomal STR profiles is the match probability, which is the conditional probability that a particular individual X has a matching profile, given that the queried contributor Q has it.[3] Matching at a single, autosomal STR allele is relatively common: typically a few percent of individuals from the same population share a given allele. The probability of matching is increased when X is a relative of Q, but for typical population sizes most of the individuals sharing a given allele are not closely related to Q.

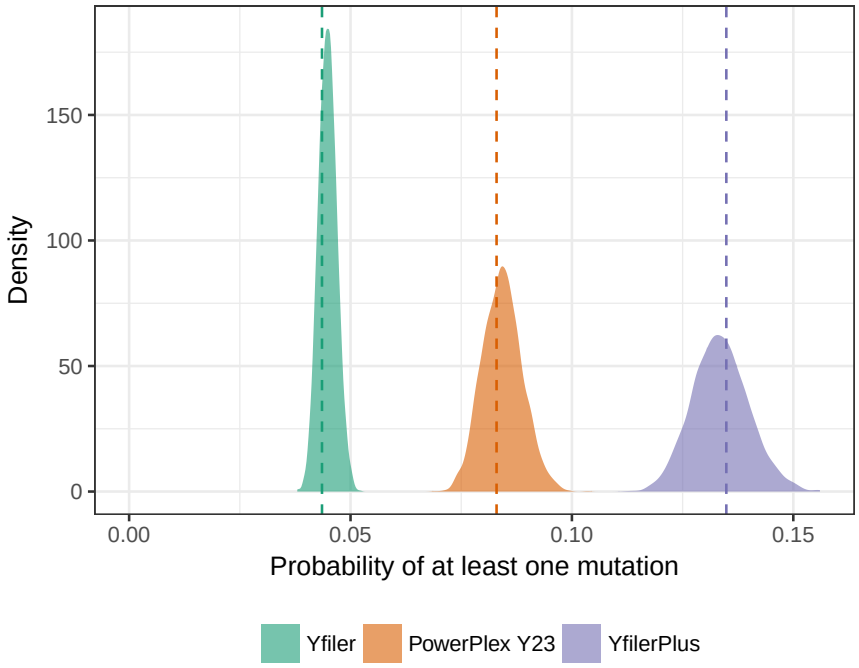

Figure 1: The posterior distributions of the whole-profile per-generation mutation rates for each of three profiling kits: Yfiler, PowerPlex Y23, and Yfiler Plus (see Materials and Methods. Vertical dashed lines indicate maximum-likelihood estimates. See Table S1 for details of the underlying mutation count data.

Because of the absence of recombination, the situation for Y-STR profiles is different.[4] The haplotype mutation rate is approximately the sum of the mutation rates at each STR locus, which can be above 0.1 per generation for profiling kits in current use (Fig. 1). This, together with the large space of possible haplotypes, implies that the probability of a matching Y profile between distantly-related males is negligible,[5] and in practice matches only occur between males connected by a relatively short paternal lineage path. However, this lineage path can include many tens of meioses, so that Q can have many such relatives and the relatedness will typically be unrecognised. Thus an attempt to exclude known patrilineal relatives of Q as alternative sources of the DNA will typically be futile. Similarly, while matching surnames can be a guide to matching Y profiles in many societies,[6] the correlation

may not be strong enough for this to be useful in eliminating alternative sources.

The number of matching patrilineal relatives of Q is insensitive to population size, and the latter is difficult to define in forensic contexts. These facts make it difficult to develop an approach to weight-of-evidence based on population match probabilities that is relevant to a specific crime. Moreover, in most approaches to weight-of-evidence for Y profiles, a count of the profile in a database plays a central role,[7–10] but because of (i) the large number of distinct Y haplotypes, (ii) the concentration of matching haplotypes in extended patrilineal clans that may be clustered geographically or socially,[11] and (iii) the fact that the databases are not scientific random samples,[12] database information is of limited value, and difficult to interpret in a valid and useful way.[2,5]

For these reasons we believe that current approaches to evaluating Y profile evidence are unsatisfactory. Proposals have been made that use population genetic models to allow for coancestry,[13] but the problem remains of setting the coancestry parameter. The appropriate value depends on the fraction of patrilineal relatives of Q in the relevantpopulation, which cannot be approximated from published data. Coalescent-based modelling of the match probability can make full use of the haplotype and database information,[14] but is computationally demanding and requires the database to be a random sample from the relevant population. Because of the former limitation, to our knowledge it has not yet been used in actual court cases.

Instead of a population fraction or match probability, we focus on the number of males with Y profile matching that of Q, and we propose a simulation model, implemented in easy-to-use software, to approximate its distribution (Fig. 2). Key parameters of the model include the locus mutation rates, and the standard deviation in a man's number of male offspring (SDO). We extend our simulation framework to provide a novel and powerful approach to using counts from a database assumed to be sampled randomly in the relevant population. While the random sampling assumption may be unrealistic, it serves to illustrate the limited value of database counts even in this optimistic setting.

Our software and method allow for the first time Y profile evidence to be quantified in a way that is valid and directly interpretable to courts.

2 Methods

2.1 Marker sets and mutation

We consider three Y-STR profiling kits: Yfiler (17 loci), PowerPlex Y23 (23 loci), and Yfiler Plus (27 loci). All markers from Yfiler are present in the other two kits, but PowerPlex Y23 has two markers not present in Yfiler Plus (DYS549 and DYS643) and Yfiler Plus has six markers not present in PowerPlex Y23 (DYS627, DYS460, DYS518, DYS449 and two markers at DYF387S1).

All three kits include both copies of the duplicated locus DYS385. In our simulation study, DYS385a and DYS385b are treated as two independent loci, whereas in practice the order of the duplicated locus is unknown and so e.g. allele pair 13/14 cannot be distinguished from

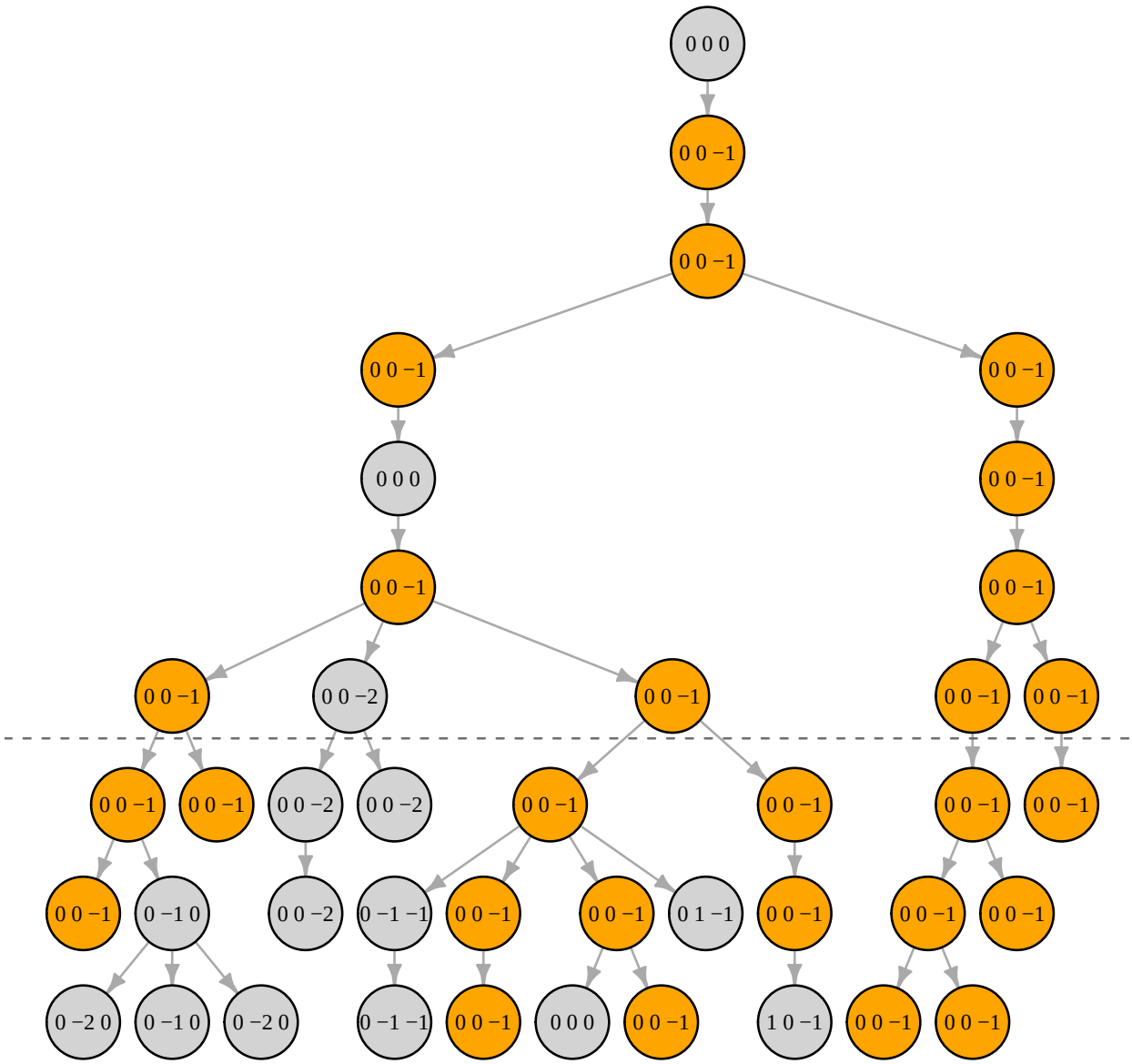

Figure 2: Simplified illustration of the process of counting haplotype matches (here, 3-locus haplotypes simulated over 10 generations with mutation rate 0.1 per locus per generation). Below the dashed line circles indicate males in the final three generations, who we label "live". A live male Q was sampled and observed to have haplotype $(0, 0, -1)$. The figure shows the founder that is ancestral to Q and all descendants of that founder, highlighting in orange those with haplotype $(0, 0, -1)$. In this instance there are 16 live males with haplotype $(0, 0, -1)$; any of these 16 could be Q – the resulting figure would be the same.

14/13. We use the same mutation rate for DYS385a and DYS385b, estimated by halving the reported mutation count for locus DYS385. Yfiler Plus includes another duplicated locus, DYF387S1, for which similar comments apply.

Mutation count data for these loci are given in Table S1. To allow for uncertainty in mutation rates, we assume a Beta$(1.5, 200)$ prior distribution (Fig. S1), so that the posterior distribution given the count data is Beta$(x + 1.5, y + 200)$, where x and y are the numbers of meioses in which mutations did and did not occur for that locus.

2.2 Population and mutation simulations

We adopt a Wright-Fisher model with a constant per-generation population size N. Variable N over time could be incorporated in the simulation. Matching individuals are usually linked by relatively recent common ancestors, and over such limited time intervals the constant N assumption will often be adequate, but perhaps not for populations that have grown rapidly in the past 200 years.

In practice, because the haplotype mutation rate is high, $|\Omega|$, the number of live males with Y profile matching that of Q, is usually at most several tens, and provided that $N > 10^3$, it is insensitive to N. Although larger than necessary, we chose $N = 10^5$ to generate a range of pedigree structures for multiple sampling of Q. We generate haplotypes for the live individuals by assigning allele 0 at all markers for each founder, and running a neutral, symmetric, single-step mutation process forward in time over $G = 250$ generations, so that with high probability all lines of descent from founder to current generation include multiple mutations. The mutation model is a reasonable first approximation to the STR mutation process[15,16] and adequate here because most of the matches of interest are between individuals sharing the same haplotype with no intervening mutations, in which case only the mutation rate is important, not any other details of the mutation process. However, some matches arise with intervening mutations that cancel, and so deficiencies in this model can have a small impact on our results.

The results presented here are based on 10^6 distinct values of $|\Omega|$: for each profiling kit and value of SDO, we repeated the population simulation 10 times, and for each of these we simulated the mutation process 100 times, each time starting by resampling each mutation rate from its posterior distribution. For each population/mutation simulation we made 10^3 selections of Q, each time recording the number of his Y-profile matches.

2.3 Comparison of simulated and real databases

The available Y-profile databases are not random samples from a well-defined population, and may over- or under-sample patrilineal clans. We are unable to mimic database selection processes, which are diverse, and the populations from which real databases are sampled are often only loosely defined. Therefore we cannot rigorously test the realism of our simulations against empirical data, but some comparison is informative.

We compared results from our population and mutation simulations against data from 6 databases of PowerPlex Y23 profiles drawn from: Central Europe (n = 5,361), Eastern

Europe ($n = 1{,}665$), Northern Europe ($n = 903$), Southeastern Europe ($n = 758$), Southern Europe ($n = 1{,}462$) and Western Europe ($n = 2{,}590$).[17] For each real database, we randomly sampled databases with the same n from simulated populations with both $N = 10^5$ and $N = 10^6$. For each profiling kit and value of N, we fixed SDO = 1.1 and repeated the population simulation 10 times, for each population we simulated the mutation process 10 times, and for each of these we simulated 10 databases. Hence, each boxplot in Fig. 3 is based on 1,000 datapoints.

2.4 Identifying matching males

Given a population and mutation simulation, we treat the males in the final three generations as potentially of interest, for brevity we label them as "live". We choose one at random among the live males to represent Q, the queried source of the Y profile and then identify the set Ω of live males carrying the same Y profile as Q. The males in Ω represent the set of potential sources of the DNA of interest to a court faced with deciding whether or not Q was the true source. Depending on the case circumstances, some males with Y profiles matching Q will be of an age or live in a location that makes them unlikely to be a source. We do not attempt to model such information and believe that the court is better placed to make such judgments than a forensic DNA expert.

2.5 Standard deviation in number of male offspring

Competition for mates and other factors can lead to a high variance in reproductive success among males in many non-human species and some human societies, notably those in which polygyny is practised.[18] Our simulation model allows us to vary SDO, the standard deviation in the number of male offspring of each male. We can thus investigate the extent to which higher SDO leads to more matches with the Y profile of Q.

We posit a symmetric Dirichlet distribution that specifies the probability for each man in a generation to be the father of an arbitrary male in the next generation. The standard Wright-Fisher model assigns all N men probability $1/N$ to be the father, which is the limiting case as the Dirichlet parameter α increases and implies SDO = 1. Our main results have used $\alpha = 5$, which implies that a man's relative probability to be the father has 95% probability to lie between 0.32 and 2.05, compared with a constant one under the standard Wright-Fisher model. For comparison, we also consider $\alpha = 1$, when each man's relative paternity probability has an Exponential(1) distribution, with 95% equal-tailed interval from 0.025 to 3.7.

To derive the SDO corresponding to an α value, note that under our model a man's total number of male children is, to a close approximation when N is large, Poisson distributed. The mean is specified by his relative probability to be the father of a given male in the next generation, which is Gamma distributed with mean 1. If $X \sim \text{Gamma}(\alpha, \beta)$ (shape α, scale β, mean $= \alpha/\beta$) and $Y \mid X = x \sim \text{Poi}(x)$, then Y has a Gamma-Poisson distribution with standard deviation $\beta\sqrt{\alpha(1+1/\beta)}$. We have $\beta = 1/\alpha$ (since mean = 1) and so SDO = $\sqrt{1+1/\alpha}$. For $\alpha = 1$, SDO $= \sqrt{2} \approx 1.4$ and when $\alpha = 5$, SDO $= \sqrt{1.2} \approx 1.1$.

2.6 Importance weighting to condition on database match count

A low database count of the profile suggests a low population count. However because in a large population almost all Y profiles are very rare, a profile count from a database of modest size (typically in practice up to a few thousand) provides little information.[5] Worse, Y-STR databases are not random samples, and may be affected by sampling biases. For example, attempts are often made to remove known relatives from databases, whereas (possibly remote) patrilineal relatives are the most plausible alternative sources of matching profiles.

Each simulated value of $|\Omega|$ is drawn from a distribution that is not conditional on a database match count for the profile. To gain some quantitative interpretation of an observation of m copies of the profile of Q in a database of size n, we assume that the database has been sampled randomly from the live males. Although this assumption is unrealistic, it can be useful for illustration.

Given a count of m observed in a database of size n, we seek to modify the distribution $|\Omega|$ to take this information into account. We could have achieved this in our simulation framework as follows: for every simulation of Q and Ω, we sample a database of size n at random from the live males, and if the number of males from Ω included in the database equals m, we retain the simulation, otherwise we reject it. However, this rejection sampling approach can be inefficient, particularly when $m > 1$.

Instead we use importance sampling to reweight the $|\Omega|$ values to reflect conditioning on m. We assume that the database represents a random sample of size n from the $3N$ live individuals and write $\pi = |\Omega|/(3N)$ for the fraction of the live individuals that match the profile of Q. The importance-sampling weights are the binomial probabilities of observing m copies of the profile in the database of size n:

$$\binom{n}{m}\pi^m(1-\pi)^{n-m}. \tag{1}$$

after normalising to have an average of one over the full simulation.[19] Intuitively, a large value of m is implausible if $|\Omega|$ is small, and so the corresponding importance-sampling weight should be small.

To assess the efficiency of the importance sampling, the effective sample size (ESS) was calculated[19] for $n = 100$ and 1,000 and for m from 0 to 6. The results (Fig. S2), show that the ESS decreases rapidly as m increases, so that the approximation to the conditional distribution of $|\Omega|$ is typically poor for $m > 3$.

There may be one or more profiles that are sufficiently common in the database that the count m can be directly interpreted. For example, if $m/n = 0.05$ say, and n is large, say $n > 10^3$, a direct interpretation is that about 5% of the population carries this profile. Our simulations suggest that for realistic parameter values and modern profiling kits with high mutation rates, it is unlikely for any profile to reach such a high relative frequency in a large population. A high observed value of m may be due to a patrilineal clan being over-represented in the database. Even for common profiles, it remains true that almost all

of those with the profile descend from a relatively recent patrilineal ancestor, and so they are unlikely to be well mixed in the population relevant to a given crime.

3 Results

3.1 Comparison of simulated and real databases

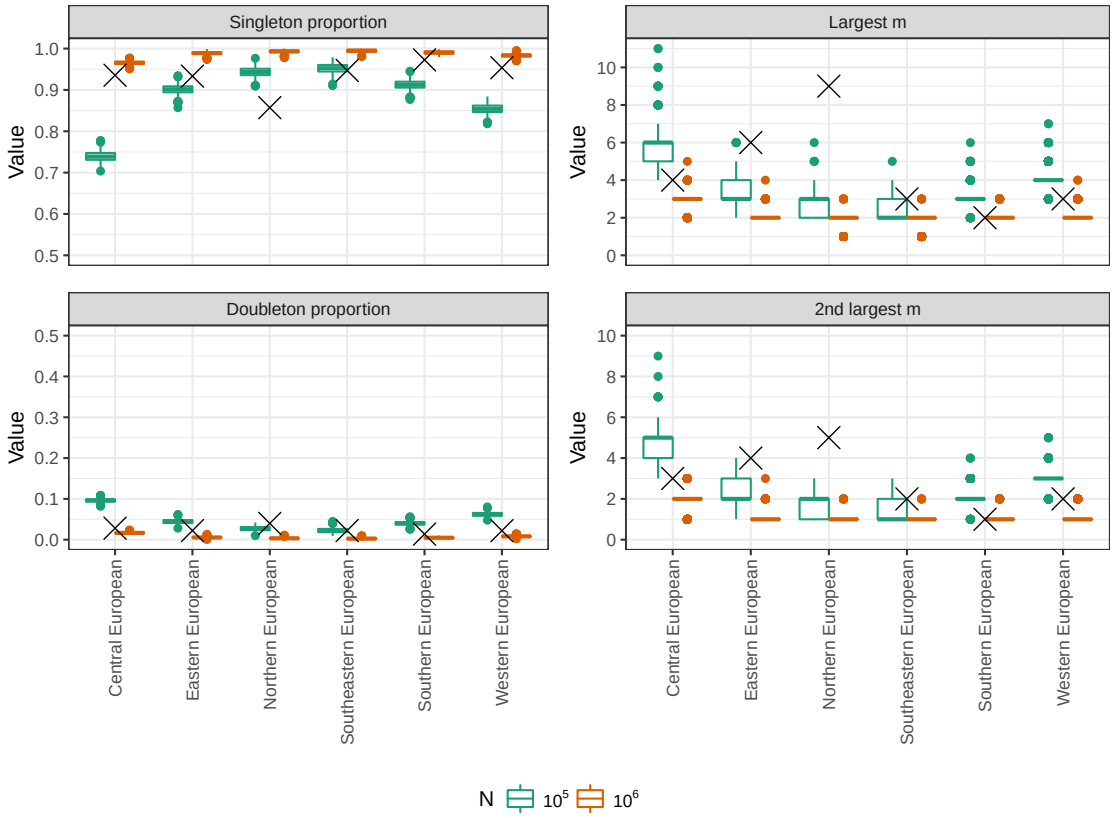

Figure 3: Properties of PowerPlex Y23 profile databases drawn at random from simulated populations with $N = 10^5$ (green boxplots) and $N = 10^6$ (red boxplots) and real databases of the same sizes (crosses).[17] The left panels show the fractions of singleton profiles and doubletons (profiles arising exactly twice). The right panels show the counts of the two most common profiles.

We compared six real Y profile databases with databases of the same sizes sampled from simulated populations of sizes $N = 10^5$ and $N = 10^6$. Fig. 3 shows a broad similarity between real and simulated databases, with properties of real databases often lying between the $N = 10^5$ and $N = 10^6$ simulated databases. However these values of N are much lower than the census male populations corresponding to the database labels, suggesting restricted

sampling. Northern Europe is an outlier, which could reflect lower effective population size or non-random sampling, or both. Almost half the Northern Europe database is drawn from Finland, which includes in the north of the country one of the most important population isolates in Europe. The two most common profiles were both restricted to Finnish samples.

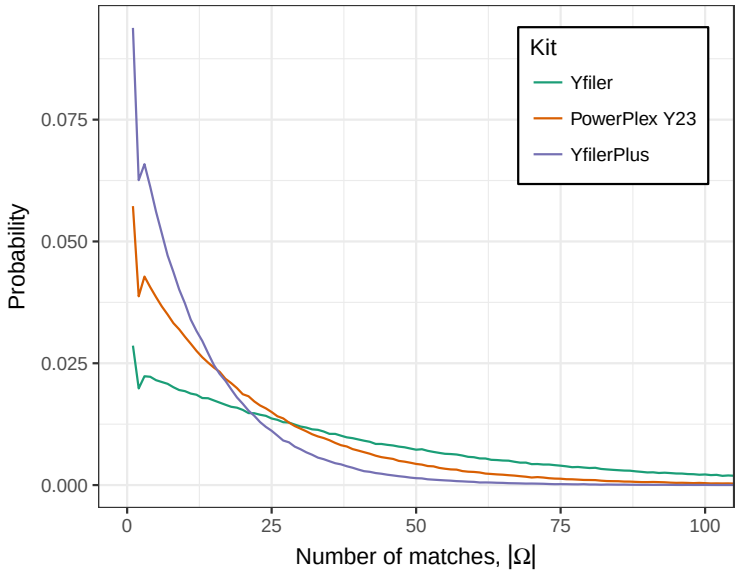

Figure 4: The distribution of $|\Omega|$, the number of live males with Y profile matching that of Q, for each of three profiling kits when the standard deviation in number of male offspring (SDO) is 1.1. See Table 1 for numerical properties of these unconditional distributions and conditional distributions of $|\Omega|$ given database count information.

3.2 The number of matching Y profiles

Fig. 4 shows, for each of the three profiling kits, the distribution of $|\Omega|$, the number of live males with profile matching that of Q. Here "live" means in the final three generations of the population simulation. As expected, $|\Omega|$ tends to be larger when there are fewer loci and so a lower haplotype mutation rate. However, the distributions are highly dispersed and overlap substantially. If Q is a defendant who denies being a source of the DNA, larger values of $|\Omega|$ are more helpful to his case. Therefore if a court wishes to consider scenarios that are helpful to the defendant while remaining realistic, it might regard the 95% or 99% quantile as a useful summary of the distribution of $|\Omega|$ (Table 1, first row).

3.3 Effect of standard deviation of offspring number

Our results so far have used SDO = 1.10, which compares with SDO = 1.07 estimated for the modern US population.[20] Table 2 summarises the effect of varying SDO on unconditional

m	n	Median	Mode	95% quantile	99% quantile
Unconditional	-	9	1	37	57
0	100	9	1	37	57
0	1,000	9	1	36	55
0	10,000	6	1	26	40
1	100	21	13	58	81
1	1,000	20	13	56	79
1	10,000	15	8	41	58
2	100	33	25	77	103
2	1,000	32	25	74	99
2	10,000	23	17	55	73

Table 1: Distribution of the number of males with a matching Yfiler Plus profile with and without database count information. Row 1 gives key properties of the Yfiler Plus distribution shown in Fig. 4. Other rows show corresponding properties of conditional distributions given an observation of m copies of the profile in a randomly-sampled database of size n. Here, SDO = 1.1. Corresponding values for the other profiling kits and for different SDO are given in Table 2 (unconditional) and Tables S2, S3 and S4 (conditional on database information).

distributions of $|\Omega|$. In addition to the standard Wright-Fisher value SDO = 1, we consider SDO = 1.4, which is very high for a modern developed society but higher values have been recorded in human societies.[18] More detailed results are given in Tables S2, S3 and S4, and Fig. S3.

3.4 Role of database information

Table 1 gives properties of the conditional distribution of $|\Omega|$ given three values of the database count m, for each of three database sizes n, drawn from a simulated population with $N = 10^5$ per generation. As expected, $|\Omega|$ increases stochastically with m, but the distributions substantially overlap and the practical impact of the dependence on m for the decision process of a court may be only modest. For example, with the most discriminatory kit that we consider, Yfiler Plus, the upper 99% quantile of the unconditional distribution of $|\Omega|$ is 57. If, however, we note that the profile of Q is unobserved in a database of size $n = 10^3$, this apparently useful information only has the effect of reducing the 99% quantile from 57 to 55, even under optimistic assumptions about the database. In reality, the males in Ω are likely to be clustered geographically and/or socially, and the database is usually not a random sample from the population, further reducing the value of database information.

Kit	SDO	Median	Mode	95% quantile	99% quantile
Yfiler	1	24	1	101	155
Yfiler	1.1	29	1	121	186
Yfiler	1.4	46	1	199	304
PowerPlex Y23	1	13	1	52	80
PowerPlex Y23	1.1	15	1	62	95
PowerPlex Y23	1.4	23	1	101	155
YfilerPlus	1	8	1	32	48
YfilerPlus	1.1	9	1	37	57
YfilerPlus	1.4	13	1	59	92

Table 2: The effect of varying the profiling kit and SDO. Properties of the unconditional distribution of $|\Omega|$, the number of live individuals with profile matching that of Q, for each of the three kits, and for each of three values of SDO (=). "Unconditional" means that database profile counts (m) are not taken into account. The penultimate row is the same as the first row of Table 1.

4 Discussion

4.1 Presentation in court

Based on the results from simulations such as those underlying Fig. 4, evidence could be presented at court along the following lines:

> "A Y-chromosome profile was recovered from the crime scene. Mr Q has a matching Y profile and so is not excluded as a contributor of DNA. Using population genetics theory and data, we predict (with 99% probability) that the number of males in the population with a matching Y profile is at most 57. These men or boys span a wide range of ages and we don't know where they live. They are all paternal-line relatives of Q, but the relationship may extend over many father-son pairs, well beyond the known relatives of Q. Since these individuals share paternal-line ancestry with Q, they could also be similar to Q in ethnic identity, language, religion, physical appearance and place of residence."

If there is database frequency information, the court could be further advised, for example:

> "The Y profile of Q was not observed in a database of 1,000 profiles. Because the database does not represent a scientific random sample and because paternal-line relatives may tend to be clustered in geographic and social groups that are not well sampled in the database, it is difficult to interpret this information. If the database were a random sample from the population, its effect would be to reduce the 99% upper limit on the number of matching males from 57 to 55."

Depending on the circumstances of the case, a judge might further instruct members of the jury:

> "If you consider that there may be around 55 males of different ages with a Y profile matching that of Q, and that these males may tend to resemble Q in some characteristics more than random members of the population, your task is to decide whether all the evidence that has been presented to you is enough to convince you that Q is the source of the crime scene DNA, and not one of these other males with the same Y profile."

It should be clear from these considerations that a matching Y profile, taken alone, can never suffice to establish convincingly that Q is a source of the crime-scene DNA. However, it remains very powerful evidence that can reduce the number of alternative sources from perhaps several millions, for a crime in a large city, down to just a few tens. If an autosomal DNA profile is also available that includes Q then the remaining task for the prosecution of establishing that Q is the source, rather than one of the 55 or so potential Y-matchers, will usually be readily achievable even if the autosomal profile is complex, for example due to DNA from multiple sources. Alternatively, non-DNA evidence suffice to complete the task of convincing a jury that Q is the source of the Y profile.

We have here only considered DNA samples with a single male contributor. A similar approach can be applied for multiple male contributors, see Section A in Supplementary Material, Fig. S4 and Table S5.

4.2 Coancestry, kinship and sampling adjustments

For autosomal DNA profiles a coancestry adjustment is generally recommended to allow for remote shared ancestry between Q and X, the hypothetical alternative source of the DNA. Similarly if Q and X might be close relatives, a match probability can be computed using explicit kinship parameters. Sampling adjustments are also sometimes recommended, to allow for the finite database sample size. None of these adjustments is needed for the method proposed here, because all matches with related individuals are modelled in the simulations, as is the size of the database.

4.3 Likelihood ratio reporting

Forensic weight-of-evidence is often best quantified using a ratio of likelihoods under prosecution and defence hypotheses, which in simple settings reduces to a match probability. We propose a different approach, reporting to the court an estimate of the number of males with matching Y profiles, but we do not believe that there is any contradiction or conflict. Indeed, in the mid-1990s when autosomal DNA profile match probabilities were not as minuscule as they have become, the England and Wales Court of Appeal[21] recommended that, instead of a match probability, courts be informed of the expected number of matching individuals in a relevant population. This recommendation was followed for some time, but match probabilities became so small that the expected number of matching individuals was a small fraction,

which was deemed potentially confusing to jurors. More importantly, for complex profiles the concept of the number of matching individuals is problematic, because the "match" may only be partial.

If the size of the population of possible sources of the DNA can be identified as N' and if the number of males with matching Y profiles is assumed to be k, then a likelihood ratio can be reported as N'/k. However because of the complications that we have discussed we don't support the likelihood-ratio formalism for Y profile evidence, even though we do support its use for many other forms of forensic evidence.

4.4 Mitochondrial DNA profiles

Similar issues arise for mitochondrial DNA (mtDNA) profiles.[22] The mutation rate for the entire mtDNA genome is an order of magnitude lower than for current Y profiling kits, and hence the sets of matching individuals tend to be much larger. This makes the problems with non-random databases less severe, though still important. Despite their lack of discriminatory power relative to autosomal or even Y profiles, mtDNA profiles can be invaluable when matrilineal relatedness is of interest or when nuclear DNA is unavailable, for example in old bones, teeth or hair shafts, or for some highly-degraded samples.

4.5 Software

Our open source (Apache License) R package malan (MAle Lineage ANlysis) available at `https://github.com/mikldk/malan` performed the analyses described in this paper. We also provide an online demonstration app based on malan, available at `http://apps.math.aau.dk/ychr-matches/`. Mutation count data are provided for the Yfiler, PowerPlex 23 and Yfiler Plus kits.

4.6 Acknowledgments

The authors wish to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge UK, for support and hospitality during the programme Probability and Statistics in Forensic Science, where this paper was conceived. The programme was supported by EPSRC grant no EP/K032208/1. We thank François Balloux (UCL), Poul Svante Eriksen (AAU) and Niels Morling (UCPH) for helpful discussions.

References

[1] JM Butler, AE Decker, MC Kline, and PM Vallone. Chromosomal duplications along the Y chromosome and their potential impact on Y-STR interpretation. *J Forensic Sci*, 50(4):853–9, 2005.

[2] P de Knijff. Son, give up your gun: Presenting Y-STR results in court. *Profiles in DNA*, 6(2):3–5, 2003.

[3] C.D. Steele and D. Balding. *Weight of evidence for forensic DNA profiles*. Wiley, 2nd ed. edition, 2015.

[4] A Caliebe, A Jochens, S Willuweit, L Roewer, and M Krawczak. No shortcut solution to the problem of Y-STR match probability calculation. *Forensic Science International: Genetics*, 15:69–75, 2015.

[5] C Brenner. Understanding Y haplotype matching probability. *Forensic Science International: Genetics*, 8:233–43, 2014.

[6] T King and M Jobling. What's in a name? y chromosomes, surnames and the genetic genealogy revolution. *Trends in Genetics*, 25(8):351–60, 2009.

[7] Charles H. Brenner. Fundamental problem of forensic mathematics – The evidential value of a rare haplotype. *Forensic Science International: Genetics*, 4(5):281–291, 2010.

[8] 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*, 7:264–271, 2013.

[9] Mikkel Meyer Andersen, Poul Svante Eriksen, and Niels Morling. The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. *Journal of Theoretical Biology*, 329:39–51, 2013.

[10] Giulia Cereda. Impact of model choice on LR assessment in case of rare haplotype match (frequentist approach). *Scandinavian Journal of Statistics*, (to appear), 2017.

[11] L Roewer. Y chromosome STR typing in crime casework. *Forensic Sci Med Pathol*, 5:77–84, 2009.

[12] P Gill, C Brenner, B Brinkmann, B Budowle, A Carracedo, and M Jobling. DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Forensic Science International*, 124:5–10, 2001.

[13] J Buckleton, M Krawczak, and B Weir. The interpretation of lineage markers in forensic DNA testing. *Forensic Science International: Genetics*, 5(78-83), 2011.

[14] I Wilson, M Weale, and D Balding. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J Roy Statist Soc A*, 166(2):155–188, 2003.

[15] F Hao and J Chu. A brief review of short tandem repeat mutation. *Genomics, Proteomics & Bioinformatics*, 5(1):7–14, 2007.

[16] K Ballantyne, M Goedbloed, R Fang, O Schaap, O Lao, A Wollstein, Y Choi, K van Duijn, M Vermeulen, S Brauer, M Furtado, and M Kayser. Mutability of Y-chromosomal microsatellites: Rates, characteristics, molecular bases, and forensic implications. *The American Journal of Human Genetics*, 87:341–353, 2010.

[17] J Purps, S Siegert, S Willuweita, M Nagya, C Alvesc, R Salazar, and S Angustia. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International: Genetics*, 12:12–23, 2014.

[18] GR Brown, KN Laland, and MB Mulder. Bateman's principles and human sex roles. *Trends Ecol Evol*, 24(6-14):297–304, 2009.

[19] Augustine Kong. A Note on Importance Sampling using Standardized Weights. Technical Report 348, Department of Statistics, University of Chicago, 1992. `https://galton.uchicago.edu/techreports/tr348.pdf`.

[20] J Weeden, MJ Abrams, MC Green, and J Sabini. Do high-status people really have fewer children? *Hum Nat*, 17:377, 2006.

[21] Regina v A. Deen. Court of Appeal Criminal Division London. *Reported in: The Times*, January 10, 1994.

[22] W Parson, L Gusmaõ, D Hares, J Irwin, W Mayr, N Morling, E Pokorak, M Prinz, A Salas, P Schneider, and T Parsons. DNA Commission of the International Society of Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Science International: Genetics*, 13:134–142, 2014.

[23] S Willuweit and L Roewer. The new Y chromosome haplotype reference database. *Forensic Science International: Genetics*, 15:43–48, 2015.

A Supplementary Text: Mixtures

Crime-scene Y profiles can reflect a mixture of DNA from multiple male sources. It may be possible to identify the profiles underlying the mixture based on differing amounts of DNA or similarity to observed profiles,[9] but often they cannot be fully deconvolved with high confidence. We also performed simulations of a mixed Y profile with DNA from two male sources. For each of the three values of α and the Yfiler Plus profiling kit, we performed 10 population simulations, for each of these 10 mutation simulations were performed, and for each of these 1000 pairs of males were chosen at random to be the contributors to the mixture. For each contributor pair, the number of included live individuals was recorded, where "included" means has a Y profile consisting only of alleles observed in the mixture.

Among the included individuals will be those with profiles exactly matching one of the reference individuals. The sizes of these two sets of individuals will be approximately independent and have the same distribution as $|\Omega|$ for a single-source Y profile described above. Fig. S4 and Table S5 show that the number of included males who do not exactly match either reference individual is stochastically smaller than $|\Omega|$. Therefore the distribution of the total number of included males is conservatively estimated as the sum of three independent versions of $|\Omega|$. Further work is required to develop procedures for presenting evidence in court based on mixed Y profiles that takes available database information into account, but our simulations will provide helpful guidance.

B Supplementary figures

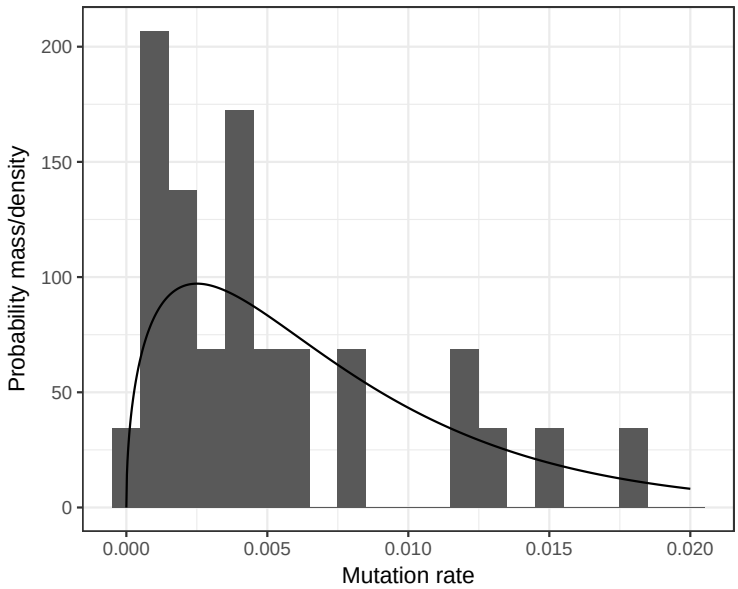

Figure S1: Histogram bars show empirical mutation rates for the 29 loci included in the three Y-STR profiling kits (rates obtained as a ratio of the counts in Table S1). The two duplicated loci (DYS385 and DYF387S1) are each represented as two loci with the same mutation rate. The curve shows the probability density for the Beta$(1.5, 200)$ prior distribution assumed for each mutation rate.

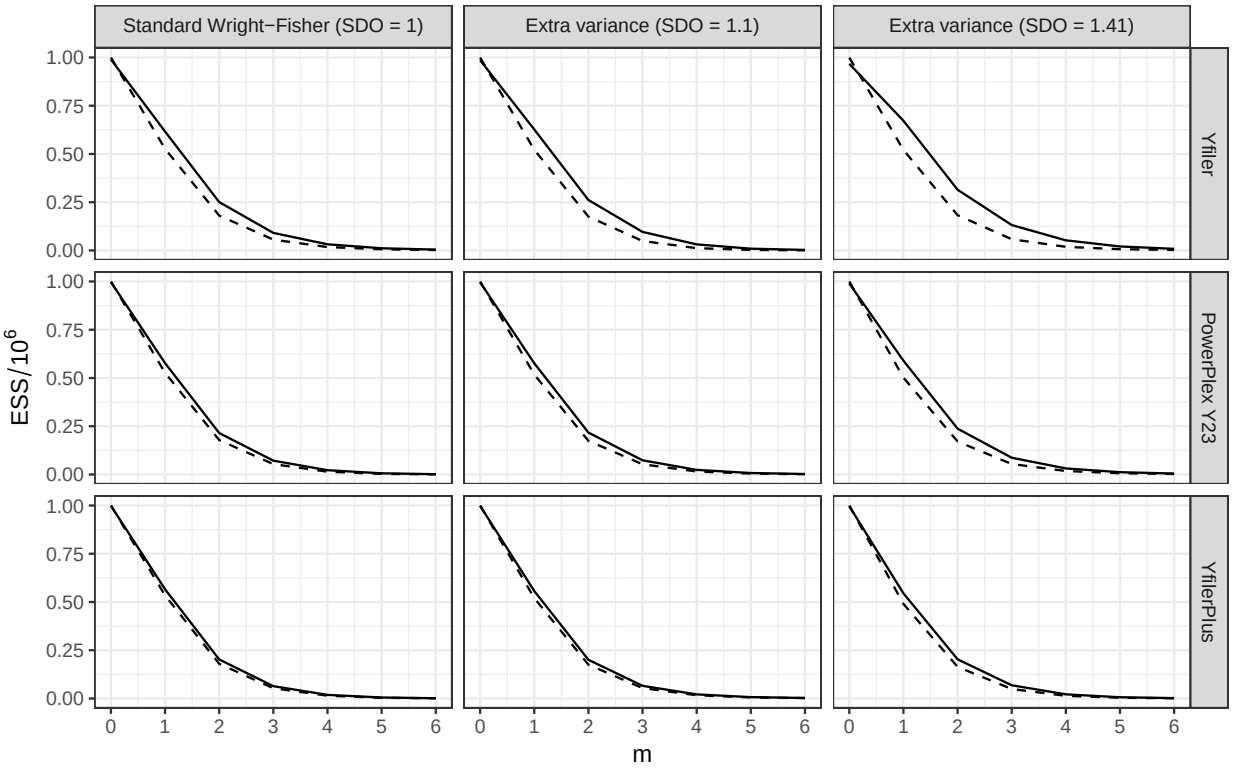

Figure S2: The effective sample size, as a fraction of the 10^6 simulated $|\Omega|$ values, for the importance sampling to approximate distributions for $|\Omega|$ conditional on database haplotype count m, for m from 0 to 6. The database sizes are $n = 100$ (dashed line) and $n = 1{,}000$ (solid line).

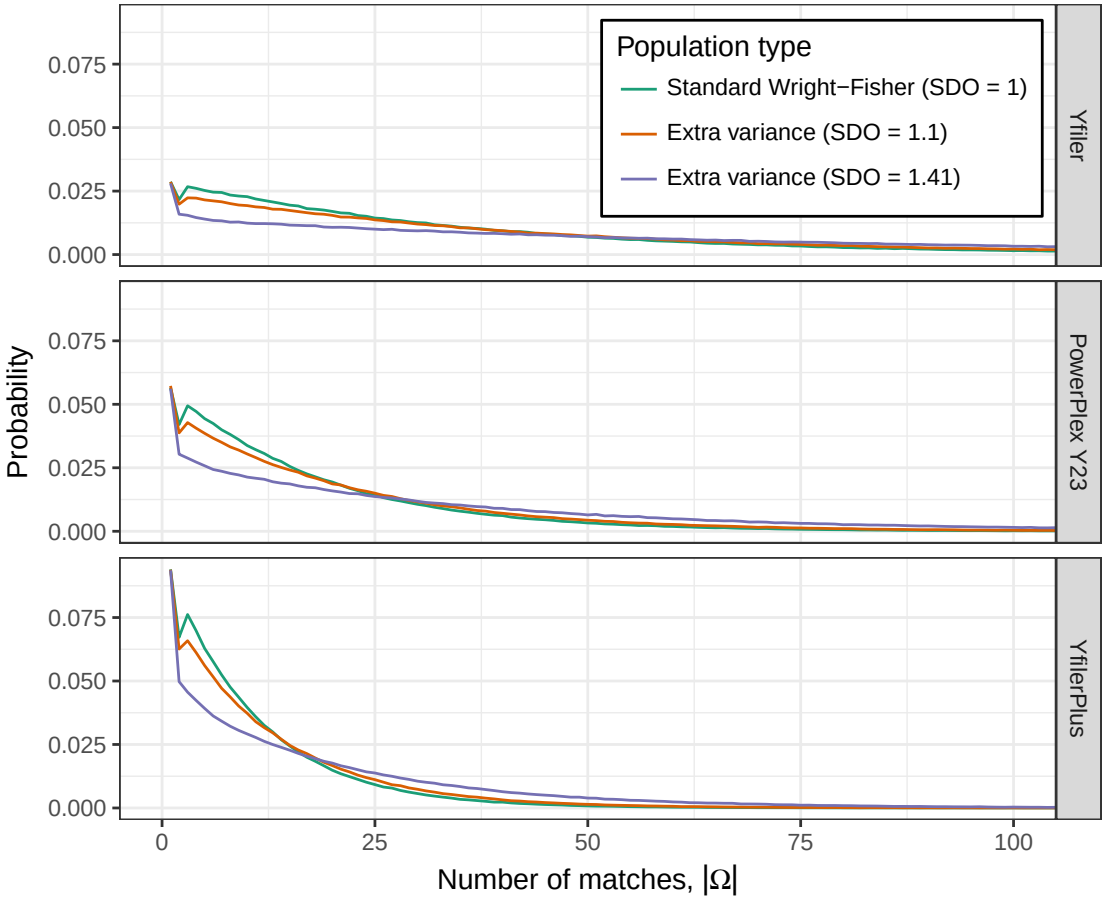

Figure S3: The distribution of $|\Omega|$, the number of live individuals with haplotypes matching that of Q, for each of the three profiling kits and three values for the standard deviation in offspring number (SDO).

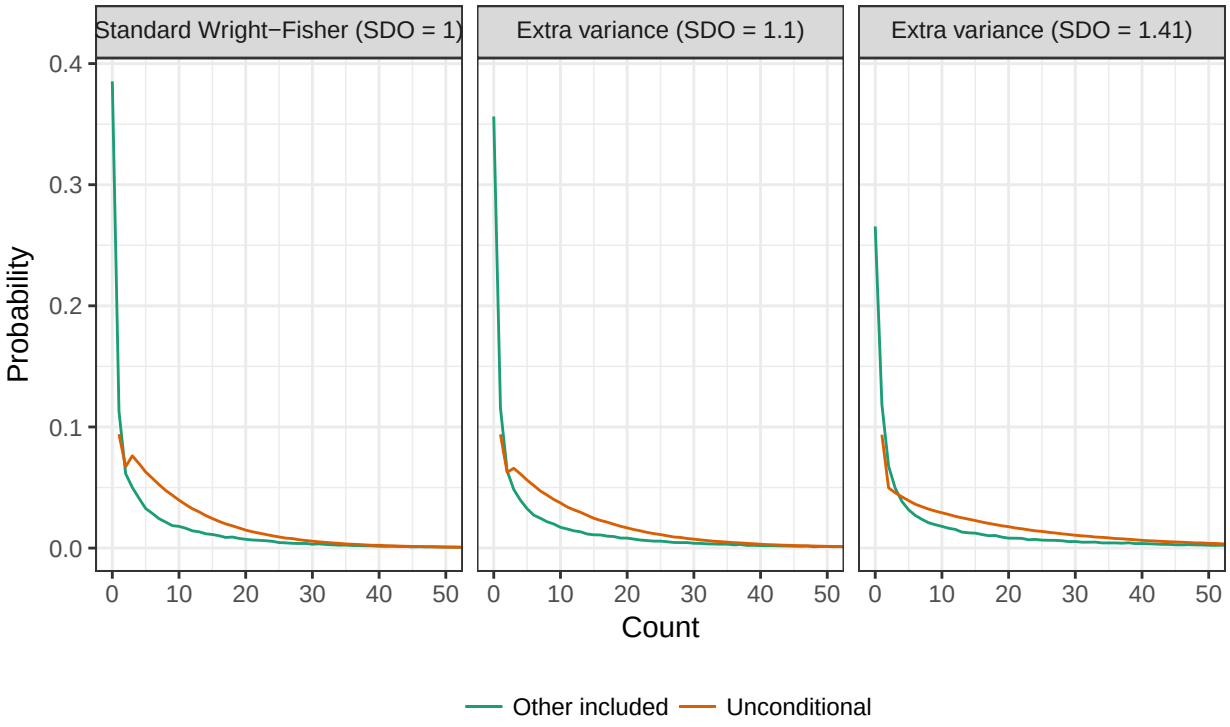

Figure S4: Distribution of the number of Y profiles included in a mixed Yfiler Plus profile arising from two male contributors. The red curve corresponds to profiles that exactly match the profile of one of the contributors. It is the same as the red ("unconditional") curve in Fig. S3 and is included again here for comparison. The green curve corresponds to "other included": males with a Y profile that consist entirely of alleles within the profiles of the two contributors, and which therefore cannot be excluded from being one of the contributing profiles, but they do not fully match either of the two contributing profiles.

C Supplementary tables

Marker	Mutations, x	Meioses, n	$10^3 \times x/n$	Yfiler	PowerPlex Y23	Yfiler Plus
DYS19	36	16,090	2.24	x	x	x
DYS389I	42	14,339	2.93	x	x	x
DYS389II	59	14,310	4.12	x	x	x
DYS390	33	15,612	2.11	x	x	x
DYS391	38	15,486	2.45	x	x	x
DYS392	8	15,418	0.52	x	x	x
DYS393	15	14,264	1.05	x	x	x
DYS385a	32	26,171	1.22	x	x	x
DYS385b	32	26,171	1.22	x	x	x
DYS439	58	10,647	5.45	x	x	x
DYS438	4	10,673	0.37	x	x	x
DYS437	13	10,652	1.22	x	x	x
DYS456	31	7,229	4.29	x	x	x
DYS458	46	7,228	6.36	x	x	x
DYS635	35	8,076	4.33	x	x	x
YGATAH4	25	8,260	3.03	x	x	x
DYS448	11	7,229	1.52	x	x	x
DYS576	32	2,174	14.72		x	x
DYS481	11	2,191	5.02		x	x
DYS549	8	2,131	3.75		x	
DYS533	8	2,177	3.67		x	x
DYS570	25	1,873	13.35		x	x
DYS643	3	2,220	1.35		x	
DYS627	21	1,766	11.89			x
DYS460	10	1,717	5.82			x
DYS518	28	1,556	17.99			x
DYS449	19	1,617	11.75			x
DYF387S1a	14	1,804	7.76			x
DYF387S1b	14	1,804	7.76			x

Table S1: Mutation count data for loci in the three kits, Yfiler (17 loci), PowerPlex Y23 (23 loci), and Yfiler Plus (27 loci). From: YHRD,[23] available at `https://yhrd.org/pages/resources/mutation_rates`, updated 17 Jan 2017. DYS385 and DYF387S1 are duplicated loci and each is treated as two independent loci (postfix 'a' and 'b'). The same mutation rate was used for DYS385a and DYS385b, estimated by halving the reported mutation count for locus DYS385, and similarly for DYF387S1.

Kit	m	n	Median	Mode	95% quantile	99% quantile
Yfiler	Unconditional	-	29	1	121	186
Yfiler	0	100	28	1	120	183
Yfiler	0	1,000	25	1	107	164
Yfiler	0	10,000	13	1	52	80
Yfiler	1	100	67	36	189	265
Yfiler	1	1,000	60	36	168	236
Yfiler	1	10,000	29	19	82	114
Yfiler	2	100	106	81	252	347
Yfiler	2	1,000	95	81	224	302
Yfiler	2	10,000	46	33	108	144
PowerPlex Y23	Unconditional	-	15	1	62	95
PowerPlex Y23	0	100	15	1	62	95
PowerPlex Y23	0	1,000	14	1	58	89
PowerPlex Y23	0	10,000	9	1	37	57
PowerPlex Y23	1	100	34	21	97	136
PowerPlex Y23	1	1,000	32	21	92	128
PowerPlex Y23	1	10,000	21	12	58	81
PowerPlex Y23	2	100	55	38	129	172
PowerPlex Y23	2	1,000	51	38	121	161
PowerPlex Y23	2	10,000	33	24	77	103
YfilerPlus	Unconditional	-	9	1	37	57
YfilerPlus	0	100	9	1	37	57
YfilerPlus	0	1,000	9	1	36	55
YfilerPlus	0	10,000	6	1	26	40
YfilerPlus	1	100	21	13	58	81
YfilerPlus	1	1,000	20	13	56	79
YfilerPlus	1	10,000	15	8	41	58
YfilerPlus	2	100	33	25	77	103
YfilerPlus	2	1,000	32	25	74	99
YfilerPlus	2	10,000	23	17	55	73

Table S2: Highlighted in grey are key properties of the three probability distributions of $|\Omega|$, the number of live individuals with haplotypes matching that of Q, shown in Fig. 4. Other rows show corresponding properties of conditional distributions given an observation of m copies of the haplotype in a randomly-sampled database of size n. Here, the standard deviation in offspring number (SDO) is 1.1. The YfilerPlus results are shown in Table 1, and repeated here for comparison with the other two profiling kits.

Kit	m	n	Median	Mode	95% quantile	99% quantile
Yfiler	Unconditional	-	46	1	199	304
Yfiler	0	100	45	1	194	298
Yfiler	0	1,000	38	1	163	250
Yfiler	0	10,000	14	1	62	95
Yfiler	1	100	109	68	306	429
Yfiler	1	1,000	92	57	257	360
Yfiler	1	10,000	35	22	99	137
Yfiler	2	100	173	131	408	547
Yfiler	2	1,000	145	131	341	457
Yfiler	2	10,000	56	42	130	174
PowerPlex Y23	Unconditional	-	23	1	101	155
PowerPlex Y23	0	100	23	1	100	154
PowerPlex Y23	0	1,000	20	1	91	139
PowerPlex Y23	0	10,000	10	1	47	73
PowerPlex Y23	1	100	56	38	159	225
PowerPlex Y23	1	1,000	51	28	144	204
PowerPlex Y23	1	10,000	27	15	75	106
PowerPlex Y23	2	100	90	67	212	282
PowerPlex Y23	2	1,000	82	56	193	257
PowerPlex Y23	2	10,000	43	33	100	133
YfilerPlus	Unconditional	-	13	1	59	92
YfilerPlus	0	100	13	1	59	91
YfilerPlus	0	1,000	12	1	56	86
YfilerPlus	0	10,000	8	1	35	55
YfilerPlus	1	100	33	20	94	131
YfilerPlus	1	1,000	31	20	89	124
YfilerPlus	1	10,000	20	11	57	80
YfilerPlus	2	100	53	37	125	168
YfilerPlus	2	1,000	50	37	118	158
YfilerPlus	2	10,000	32	25	76	101

Table S3: Similar to Table S2, but here the standard deviation in offspring number (SDO) is 1.4.

Kit	m	n	Median	Mode	95% quantile	99% quantile
Yfiler	Unconditional	-	24	1	101	155
Yfiler	0	100	24	1	100	153
Yfiler	0	1,000	22	1	91	139
Yfiler	0	10,000	12	1	48	74
Yfiler	1	100	56	31	157	220
Yfiler	1	1,000	51	31	142	199
Yfiler	1	10,000	27	16	75	105
Yfiler	2	100	88	73	209	278
Yfiler	2	1,000	80	62	188	255
Yfiler	2	10,000	42	31	100	133
PowerPlex Y23	Unconditional	-	13	1	52	80
PowerPlex Y23	0	100	13	1	52	80
PowerPlex Y23	0	1,000	12	1	50	76
PowerPlex Y23	0	10,000	8	1	34	51
PowerPlex Y23	1	100	29	20	81	114
PowerPlex Y23	1	1,000	27	14	77	108
PowerPlex Y23	1	10,000	18	12	52	73
PowerPlex Y23	2	100	46	40	108	142
PowerPlex Y23	2	1,000	44	31	103	137
PowerPlex Y23	2	10,000	29	20	69	92
YfilerPlus	Unconditional	-	8	1	32	48
YfilerPlus	0	100	8	1	32	48
YfilerPlus	0	1,000	8	1	31	47
YfilerPlus	0	10,000	6	1	24	36
YfilerPlus	1	100	17	10	49	69
YfilerPlus	1	1,000	17	10	48	66
YfilerPlus	1	10,000	13	8	37	51
YfilerPlus	2	100	28	19	65	88
YfilerPlus	2	1,000	27	19	63	85
YfilerPlus	2	10,000	21	15	49	65

Table S4: Similar to Tables S2 and S3, but assuming a standard Wright-Fisher model, which corresponds to SDO = 1 in our more general model.

Population type	Type	Median	Mode	95% quantile	99% quantile
Standard Wright-Fisher (SDO = 1)	Other included	2	0	30	53
	Unconditional	8	1	32	48
Extra variance (SDO = 1.1)	Other included	2	0	35	63
	Unconditional	9	1	37	57
Extra variance (SDO = 1.41)	Other included	3	0	56	102
	Unconditional	13	1	59	92

Table S5: Key properties of the distribution of the number of Y profiles included in a mixed Yfiler Plus profile arising from two male contributors. See Fig. S4 legend for explanations of "other included" and "unconditional".