Forensic Science International: Genetics 54 (2021) 102562

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Strategies for pairwise searches in forensic kinship analysis

Hilde Kjelgaard Brustad [a,b,*], Margherita Colucci [c], Mark A. Jobling [c], Nuala A. Sheehan [d], Thore Egeland [a]

[a] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Post box 5003 NMBU, 1432 Aas, Norway
[b] Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Post box 1122 Blindern, 0316 Oslo, Norway
[c] Department of Genetics & Genome Biology, University of Leicester, University Road, Leicester LE1 7RH, UK
[d] Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK

ARTICLE INFO

Keywords:
Kinship testing
Blind search
LR thresholds
Inbred relationships
X chromosomal markers

ABSTRACT

Testing kinship between pairs of individuals is central to a wide range of applications. We focus on cases where many tests are done jointly. Typical examples include cases where DNA profiles are available from a burial site, a plane crash or a database of convicted offenders. The task is to determine the relationships between DNA profiles or individuals. Our approach generalises previous methods and implementations in several respects. We model general, possibly inbred, pairwise relationships which is important for non-human applications and in archaeological studies of ancient inbred populations. Furthermore, we do not restrict attention to autosomal markers. Some cases, such as distinguishing between maternal and paternal half siblings, can be solved using X-chromosomal markers. When many tests are done, the risk of errors increases. We address this problem by building on the theory of multiple testing and show how optimal thresholds for tests can be determined. We point out that the likelihood ratios in a blind search may be dependent so multiple testing methods and interpretation need to account for this. In addition, we show how a Bayesian approach can be helpful. Our examples, using simulated and real data, demonstrate the practical importance of the methods and implementation is based on freely available software.

1. Introduction

Inferring the relationship between pairs of individuals is central to many forensic applications. Examples include mass fatality incidents, which can be the result of accidental catastrophes like air crashes with a list of known victims [1] or shipwrecks without passenger lists [2,3]. Other applications are natural disasters like tsunamis, where the number of victims is unknown [4] and terrorism-related events [5]. The aim is to link DNA samples from the scene to putative victims (e.g. individuals reported missing since the event) and is known as disaster victim identification (DVI). There are various other important applications like searching for relationships among individuals in mass graves of archaeological relevance [6–8]. We may also check databases collected to estimate population statistics such as allele frequencies. Duplicates and close relatives should be excluded prior to the statistical analysis to avoid biased estimates of allele frequencies [9].

As these cases involve unidentified DNA samples, a first step in the investigation is to screen the data for related samples. This initial step is referred to as a blind search [10]. It is helpful to first position the topics that we are addressing in the wider context of database searching. Assume that there is a case database of DNA profiles. This could comprise profiles obtained from a crime scene, a disaster site or a burial site. In addition, there may be a reference database of DNA profiles like a national database of convicted offenders. There are various searches that can be performed to detect pairwise relationships as illustrated in Fig. 1:
.

1. Search for duplicates, i.e., direct search, performed within or between the databases. If this is done within a database, the objective is to merge identical samples. A search between databases corresponds to the widely discussed database search problem [11].
2. Familial searching involves searching between databases [12]. A selected DNA profile is compared to the profiles of a database with the aim of detecting close kin relationships, such as parent-offspring or sibling rather than a direct match.

* Corresponding author at: Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Post box 5003 NMBU, 1432 Aas, Norway.
E-mail address: h.k.brustad@medisin.uio.no (H.K. Brustad).

https://doi.org/10.1016/j.fsigen.2021.102562
Received 30 March 2021; Received in revised form 29 June 2021; Accepted 3 July 2021
Available online 7 July 2021

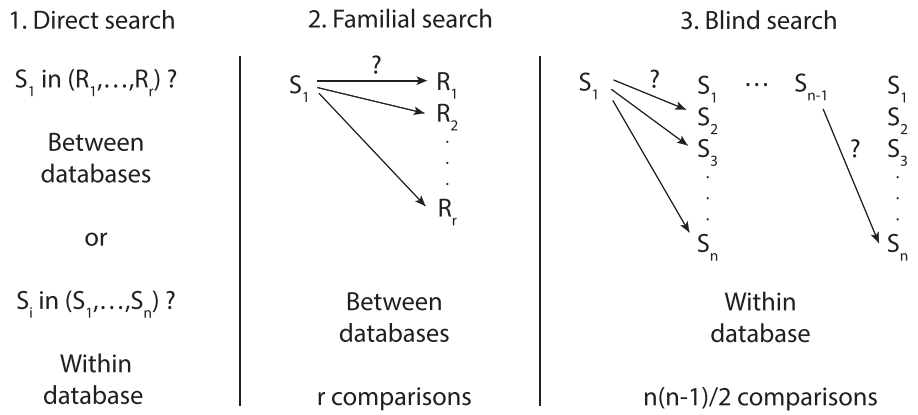

Fig. 1. Different database searches. 1. Direct search: Search for direct matches between or within databases. 2. Familial search: Search for related individuals between databases. 3. Blind search: search for related individuals within databases.

3. Blind search. This is a search within a database and is the topic of this paper.

In a blind search, comparisons are performed among all pairs of DNA samples. A likelihood ratio (LR), comparing the relationship specified by H_1 to the one specified by H_0, is computed for each pair. The LRs summarise the statistical DNA evidence. For pre-specified threshold values t_0 and t_1, small values of $LR < t_0$ are often interpreted as supporting H_0, while large values of $LR > t_1$ favour H_1. A blind search typically involves a large number of comparisons. If there are n profiles in the database, the number of comparisons is $n(n-1)/2$, e.g. 4950 comparisons for 100 profiles. The implications of this high number of pairwise comparisons in a blind search are of key concern in this paper. Also, it is not obvious how the thresholds t_0 and t_1 should be specified. Conventional thresholds used in paternity testing, for example, may not apply. The false positive rate $FPR = P(LR > t_1|H_0)$ and false negative rate $FNR = P(LR < t_0|H_1)$ should both be close to 0. Even if these error rates are small for each comparison, the probability that errors occur when many comparisons are done may be considerable. Determination of thresholds and optimisation of search strategies have been discussed in connection with database searches and familial searching [13]. The classical statistical theory of multiple testing [14] is also relevant.

Current implementations of blind search are limited to fairly simple outbred pedigree structures connecting the two individuals of interest. For example, Familias [10], a freely available kinship software package, accommodates parent offspring (PO), sibling (S), half sibling (H), first cousin (FC) and second cousin (SC) [15,16]. We model general pairwise relationship, possibly with inbreeding, using the Jacquard coefficients [17]. By including X-chromosomal markers, some additional relationships can be addressed. For instance, paternal and maternal half sisters can be distinguished.

Prior, non-DNA, information can sometimes be important. For instance, two individuals of the same age cannot possibly constitute a parent-offspring pair even if the DNA profiles suggest otherwise. To formally include prior information, we require a Bayesian approach. In the Bayesian framework we start out with a set of prior probabilities, reflecting our belief in the hypotheses, before considering any genetic data. Our belief in each hypothesis is then updated by incorporating the DNA information. Informative priors can contribute additional information to the genetic data and this will be reflected in the posterior probabilities. A more general prior distribution for pedigrees has been discussed elsewhere [18].

Our paper is structured as follows. We first review the parametric representation of relationships and the corresponding parametric likelihood and likelihood ratio, for both autosomal and X-chromosomal markers. A review of the Bayesian approach to kinship testing is given,

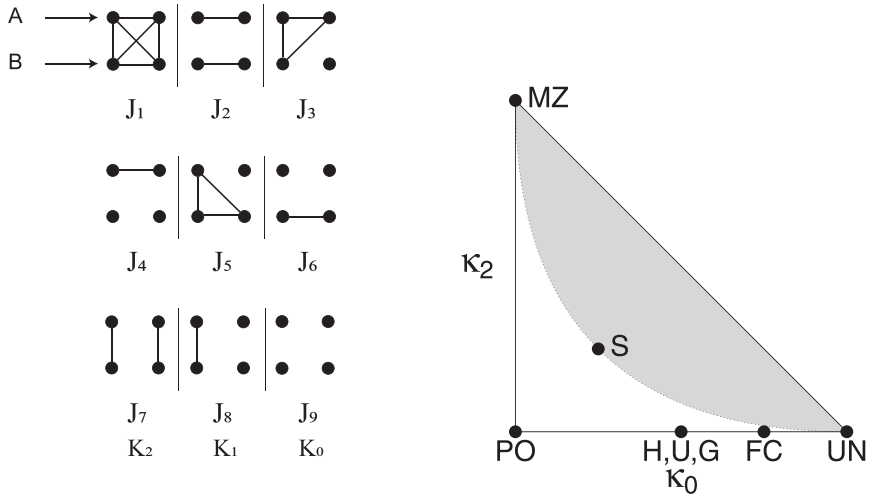

Fig. 2. Left: Jacquard states J_1, …, J_9. Dots denote alleles, and lines connect IBD alleles. Right: IBD triangle, with location of some common relationships. Abbreviations: MZ - monozygotic twins, PO - parent offspring, S - full siblings, H - half siblings, U - avuncular, G - grandparent grandchild, FC - first cousins, UN - unrelated.

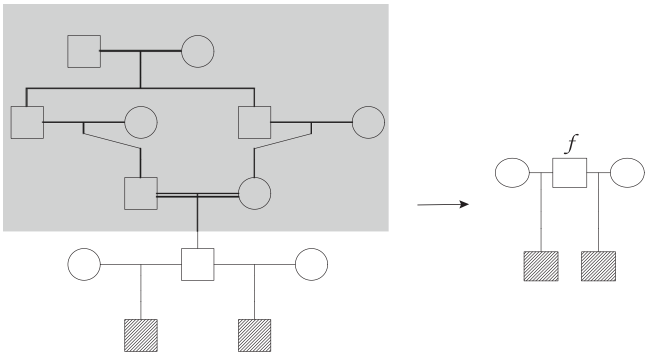

Fig. 3. Figure showing the concept of founder inbreeding, as described in Section 2.1.1. The shaded part showing the first cousin relationship, is modelled by the inbreeding coefficient f.

before we return to the likelihood ratio and its properties. These properties are incorporated when presenting the theory for evaluating the performance of a blind search. We then introduce the data used in the results section and give a brief description of our implementation. We provide several examples and conclude with a discussion of the challenges and the advantages of the work we present.

2. Review of previous results

2.1. Relatedness coefficients

Two homologous alleles are *identical by descent* (IBD) if they are identical by state (IBS) and inherited from a common ancestor. IBD is therefore defined with reference to a specified pedigree. The idea is that closely related individuals share more of their genetic material IBD than more distantly related individuals.

The simplest measure of pairwise relationships is the kinship coefficient, φ, defined as the probability that a random allele at a locus from one individual is IBD to a random allele at the same locus from another individual. This is the same as the inbreeding coefficient f of a child of these two individuals [19].

The Jacquard coefficients [17] provide a description of general pairwise relationships. The four alleles of two individuals are in one of the nine Jacquard states J_i for $i = 1, \ldots, 9$ (see left panel of Fig. 2). The probability that the alleles at a locus are in the different Jacquard states are given by the Jacquard coefficients, $\Delta = (\Delta_1, \ldots, \Delta_9)$, where $\Delta_i = \mathrm{P}(J_i)$. The coefficients sum to one.

The first six Jacquard states model inbreeding in one or both of the individuals. The only possible IBD states for two outbred individuals are J_9, J_8 and J_7, referred to as the IBD states K_0, K_1 and K_2, respectively. Thus, for two outbred individuals, the Jacquard coefficients reduce to the IBD coefficients [20], $\kappa = (\kappa_0, \kappa_1, \kappa_2)$, where $\kappa_i = \mathrm{P}(K_i)$. Since $\sum_{i=0}^{2}\kappa_i = 1$, the coefficients can be visualised in the IBD triangle, with coordinates (κ_0, κ_2). Fig. 2 shows the IBD triangle with the location of some common pedigree relationships.

Thompson [21] showed that the coefficients satisfy the inequality $\kappa_1^2 \geq 4\kappa_0\kappa_2$, which creates an inadmissible region, shown in grey in Fig. 2. This means that it is not possible to construct a pedigree connecting two outbred individuals with IBD coefficients in the inadmissible region.

2.1.1. Relatedness coefficients and founder inbreeding

By assigning a coefficient of inbreeding to one or more of the founders of a pedigree, background relatedness can be modelled [22]. Inbreeding of a pedigree founder (or several founders) affects the genetic relationship between other members of the pedigree [23], but does not necessarily make the pedigree members of interest inbred. For example, if it is suspected that two individuals are paternal half-siblings and the paternal grandparents are first cousins, as depicted in Fig. 3, the common father has an inbreeding coefficient $f = 1/16$. The IBD coefficients for these half siblings are given by $\kappa = (0.469, 0.531, 0)$ in contrast with $\kappa = (0.5, 0.5, 0)$ for the non-inbred setting with $f = 0$. It can be shown that there is some finite pedigree with founder inbreeding that corresponds to each admissible point in the IBD triangle [24].

Table 1
Likelihood of $\kappa = (\kappa_0, \kappa_1, \kappa_2)$, when observing genotypes for two individuals, for three unlinked loci, as described in the example of Section 2.3.

	P $(G\|K_0)$	P $(G\|K_1)$	P $(G\|K_2)$	L (κ)
$G_1 = (ab, ac)$	0.06	0.03	0	$0.06 \cdot \kappa_0 + 0.03 \cdot \kappa_1$
$G_2 = (bc, bb)$	0.011	0.004	0	$0.011 \cdot \kappa_0 + 0.004 \cdot \kappa_1$
$G_3 = (aa, bc)$	0.03	0	0	$0.03 \cdot \kappa_0$

2.2. The likelihood function

Our data comprise pairs of DNA profiles, genotyped at m unlinked loci, i.e., m statistically independent loci. For a single pair of individuals, A and B, let $G_j = (g_{A,j}, g_{B,j})$ denote their respective genotypes at locus j for $j = 1, \ldots, m$. The likelihood of Δ, i.e., the probability of observing the data $G = (G_1, \ldots, G_m)$ assuming Δ to be true, is

$$\mathrm{L}(\Delta) = \prod_{j=1}^{m} \sum_{i=1}^{9} \Delta_i \mathrm{P}(G_j|J_i). \tag{1}$$

The probabilities $\mathrm{P}(G_j|J_i)$ are given in Table 9 in the Appendix A. For outbred individuals, the likelihood of κ is

$$\mathrm{L}(\kappa) = \prod_{j=1}^{m} \sum_{i=0}^{2} \kappa_i \mathrm{P}(G_j|K_i), \tag{2}$$

where the probabilities $P(G_j|K_i)$ for $i = 0, 1, 2$ correspond to the last three columns of Table 9.

2.3. Parametric representation of the likelihood ratio

The likelihood ratio (LR) quantifies how much more likely it is that a set of genetic data is explained by one hypothesis H_1 than by another hypothesis H_0. In our applications, each hypothesis states a pairwise relationship, expressed by a set of relatedness coefficients Δ (or κ for outbred relationships). The LR that compares (1) for two sets of coefficients Δ_1 and Δ_0 is

$$\mathrm{LR}(\Delta_1, \Delta_0) = \frac{\mathrm{P}(G|H_1)}{\mathrm{P}(G|H_0)} = \frac{\mathrm{L}(\Delta_1)}{\mathrm{L}(\Delta_0)}. \tag{3}$$

The hypotheses H_1 and H_0 are not necessarily exhaustive, meaning that there may be other hypotheses that better explain the data.

2.3.1. Example

The purpose of this example is merely to illustrate how LRs can be easily computed for different sets of IBD coefficients using the representation in (3).

Consider two individuals genotyped at three loci. Each locus has three alleles a, b and c, with population frequencies 0.5, 0.3 and 0.2, respectively. The genotypes at each locus are given in the first column of Table 1. The likelihood of κ for each locus are given in the last column.

When comparing siblings, $\kappa_1 = (0.25, 0.5, 0.25)$ against unrelated (UN) $\kappa_0 = (1, 0, 0)$, the LR becomes

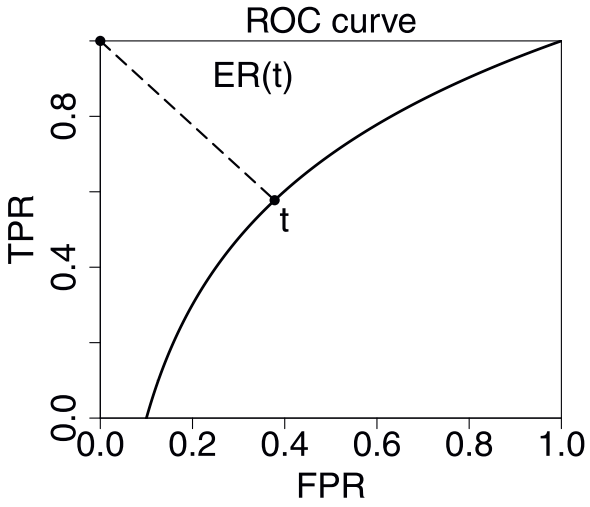

Fig. 4. Figure showing the concept of a ROC curve with a corresponding threshold. The rate TPR is plotted as a function of FPR. Each point on the curve corresponds to an LR threshold t. The dashed line shows the Euclidean distance (unweighted) from the optimal point (0,1) to the ROC curve, given by (7).

$$\begin{aligned} \mathrm{LR}(\kappa_1, \kappa_0) = & \frac{(0.06 \cdot 0.25 + 0.03 \cdot 0.5)}{0.06} \\ & \cdot \frac{(0.011 \cdot 0.25 + 0.004 \cdot 0.5)}{0.011} \\ & \cdot \frac{(0.03 \cdot 0.25)}{0.03} = 0.054. \end{aligned} \tag{4}$$

Similarly, we find the LR for half siblings (or avuncular or grandparent grandchild), $\kappa_1 = (0.5, 0.5, 0)$, against unrelated, $\kappa_0 = (1, 0, 0)$, to be 0.256. The probabilities in the middle three columns of Table 1 are independent of the tested relationships.

2.4. Properties of the LR

For specified thresholds $t_0 < t_1$, an $\mathrm{LR} < t_0$ essentially supports H_0, while an $\mathrm{LR} \geq t_1$ favours H_1. More data may be required to make a decision when $t_0 \leq \mathrm{LR} < t_1$ [25]. For simplicity, we will assume $t_0 = t_1 = t$, so that a conclusion can always be drawn.

When $\mathrm{LR} \geq t$, but H_0 is true, we have a *false positive* (FP). If $\mathrm{LR} \geq t$ and H_1 is true, we have a *true positive* (TR). We define the *false positive rate* (FPR) and the *true positive rate* (TPR) as

$$\mathrm{FPR} = \mathrm{P}(\mathrm{LR} \geq t | \mathrm{H}_0), \quad \mathrm{TPR} = \mathrm{P}(\mathrm{LR} \geq t | \mathrm{H}_1). \tag{5}$$

The TPR measures the ability to detect the relationship, while the FPR is the probability of falsely declaring a relationship. The relationship between FPR and TPR is often visualised by a receiver operating characteristic (ROC) curve [26]. Fig. 4 in Section 3.3 illustrates the concept of a ROC curve.

2.5. The Bayesian approach to kinship testing

A frequentist approach to evaluating kinship is based on the LR reflecting the probabilities of the data we have observed under two specified hypotheses. An alternative approach is provided by a Bayesian framework.

Instead of just testing one hypothesis H_1 against H_0, we consider a set of hypotheses H_i, $i = 1, \ldots, k$, each against H_0. With some prior belief in each hypothesis $\pi_0, \ldots, \pi_k$, with $\sum_{i=0}^{k} \pi_i = 1$, Bayes' theorem expresses the posterior probability of each hypothesis as

$$\mathrm{P}(\mathrm{H}_i | \mathrm{data}) = \frac{\mathrm{LR}_i \pi_i}{\sum_{j=0}^{k} \mathrm{LR}_j \pi_j}, \text{for} i = 0, \ldots, k, \tag{6}$$

where LR_i is the likelihood ratio when H_i is compared against H_0 [10]. In fact, the denominator in the LR cancels out, so (6) actually compares the likelihood of each hypothesis against all the other hypotheses jointly.

Just as for LRs, we cannot infer anything about the true relationship between the individuals as this might not be one of the hypotheses considered. For a flat prior, the posterior probabilities do not add any information to that provided by the genetic data and hence simply scale the relevant likelihoods (or LRs). More informative priors, on the other hand, can contribute additional information and this will be reflected in the posterior probabilities. For example, the three relationships half-sibling (H), avuncular (U) and grandparental (G) all have the same IBD coefficients and identical likelihoods. They are hence indistinguishable in the traditional frequentist setting and in a Bayesian setting using flat priors. Age information can easily be incorporated into the Bayesian approach and may yield different posterior probabilities.

3. Methods for blind search

3.1. The likelihood ratio for X-chromosomal markers

X chromosomal markers are increasingly used in forensic applications to supplement or replace autosomal markers for some cases of practical importance [27]. One such example is shown in Fig. 8. The females B and C are *paternal* half sibs while C and D are *maternal* half sibs. The distinction between maternal and paternal is captured by X-chromosomal markers but not by autosomal markers. The paternal half sibs share an allele IBD inherited from their father. The Jacquard coefficients and the likelihood calculation can be modified to cater for independent X-chromosomal markers (details omitted). Obviously, the sex of the individuals in the pair matters. As an example note that there are only two possibilities, or two states, for a pair of males: either they share an allele IBD or they do not.

Since the number of unlinked markers on the X chromosome is limited, linkage and linkage disequilibrium become an issue [28]. We will ignore such dependence in Section 5.5. However, relevant findings that take dependence into account can be checked using the freely available software FamLinkX [29].

3.2. Estimation of FPR and TPR

The true positive and false positive rates are determined by the hypotheses considered, number of loci, properties of each locus and the LR threshold. These rates can be calculated numerically using the algorithm described in [30]. However, this method only works for a small number of markers, say up to 10. In practice, we therefore resort to simulation. We denote estimates of FPR and TPR by $\widehat{\mathrm{FPR}}$ and $\widehat{\mathrm{TPR}}$, respectively.

Typically TPR is close to 1 and FPR close to 0. These values are generally poorly estimated from direct Monte Carlo simulation. For instance, when $\mathrm{FPR} = 0.00001$, 1 of 100000 simulations is expected to give a false positive. The conventional number of simulations in the range 100–10000 is therefore likely to return an estimate of 0. Kruijver [30] describes several methods for estimating small probabilities in forensic applications. One of these is importance sampling, which we use to estimate FPR in the results section. Details about importance sampling are given in Appendix B.

3.3. Optimal LR threshold

Intuitively we seek a threshold for LR that minimises the number of errors. Several approaches for choosing the optimal threshold have been suggested and compared [31]. We will focus on ROC-based methods. Fig. 4 shows a general ROC curve. Each point on the curve corresponds to a threshold t, with corresponding values for FPR and TPR. The value t in the figure corresponds roughly to $\mathrm{FPR} = 0.4$ and $\mathrm{TPR} = 0.6$. The upper left corner corresponds to $\mathrm{FPR} = 0$ and $\mathrm{TPR} = 1$ and is therefore

Table 2
The statistics of a blind search summarised, as described in Section 3.4. Only W_0, the number of LRs below t, and W_1, the number of LRs above t, are observed.

	Claim H_0	Claim H_1	Total
H_0 true	TN	FP	N_0
H_1 true	FN	TP	N_1
Total	W_0	W_1	N

Source:Adapted from [32].

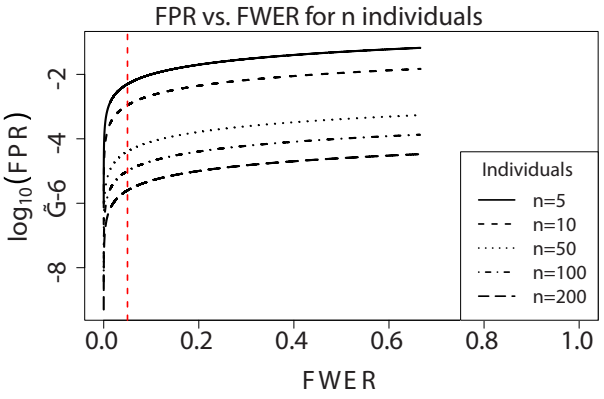

Fig. 5. Highest acceptable value of FPR, as a function of α, given by (8) in Section 3.4. Plotted for blind search with 5, 10, 50, 100 and 200 individuals. The vertical line is located at $\alpha = 0.05$.

called the optimal point. Consequently, it is reasonable to choose a threshold that minimises the weighted Euclidean distance between the ROC curve and the point (0,1),

$$\mathrm{ER}(t) = \sqrt{(w\ \mathrm{FPR}(t))^2 + (1 - \mathrm{TPR}(t))^2}, \tag{7}$$

where $w \geq 0$. In our examples the weight $w = 1$. It may be that one of the errors, typically a false positive, is more important to avoid than the other, a false negative. The relative importance of errors can be modelled by using other values of w. Because we do not know the exact values of FPR and TPR, they are replaced by their estimates.

3.4. The problem of multiple testing in blind search

When doing a blind search among n DNA profiles, we compute one LR comparing two hypotheses, H_0 and H_1, for each pair of DNA profiles, leaving us with a total of $N = n(n-1)/2$ LRs, LR_1, …, LR_N. The blind search can be repeated with a different pair of hypotheses, but here we restrict attention to a single blind search with two hypotheses H_1 and H_0 for all pairwise comparisons. If the true hypothesis is known in each case, the result of the search (or any other multiple testing scenario) can be summarised as shown in Table 2. In practice, the truth will only be known for simulated data.

Assume that the only possibilities are the relationship stated by H_1 or H_0, such that $N_0 + N_1 = N$. The number of type I errors or false positives is FP, while the number of false negatives is FN. Ideally, we want FP = 0 and FN = 0. However, this is not realistic. For a sufficiently large threshold t we will never reject H_0 and there will be no false positives, i.e. FP = 0. Similarly, there will be no false negatives, FN = 0 (which means $TP = N_1$), for a sufficiently small threshold. The challenge is to make both FP and FN acceptably small, or equivalently, make FP as small as possible and TP as close to N_1 as possible.

Even if the probability of a false positive is very small for a single pairwise comparison, the fact that there are so many tests in a blind search could lead to a substantial probability of at least one false positive. Approaches to analyse and control these false positives in a multiple testing setting have to be applied. The Family Wise Error Rate (FWER) [14] is often used for this purpose.

FWER is defined as the probability of getting at least one false positive out of N tests [32]. Let α denote the FWER. For N independent tests,

$$\alpha = \mathrm{P}(\mathrm{FP} \geq 1) = 1 - (1 - \mathrm{FPR})^N,$$

where the FPR, as defined in (5), is assumed to be the same for each test. As we illustrate in the results section, the pairwise tests in a blind search are not independent and so we use the Bonferroni bound

$$\alpha \leq N \cdot \mathrm{FPR} = \frac{n(n-1)}{2}\mathrm{FPR}.$$

Thus, to obtain an α below a given value, we choose a threshold so that

$$\mathrm{FPR}_\alpha \leq 2\alpha/(\mathrm{n}(\mathrm{n}-1)) \tag{8}$$

for a fixed sample size n. Fig. 5 plots FPR_α as a function of α, for a blind search with 5, 10, 50, 100 and 200 individuals. The red vertical line is located at $\alpha = 0.05$.

The aim is to find the threshold t that minimises ER(t) given in (7), with the constraint that $\mathrm{FPR} \leq \mathrm{FPR}_\alpha$.

4. Data and implementation

4.1. Real data and simulations

The DNA profiles evaluated in Sections 5.3 and 5.4 are from 65 individuals of Northern European origin (Germany) forming 8 pedigrees, with a variety of declared kinships up to 7th degree (the number of meioses between the persons of interest [33]). Most founders were not genotyped, and pedigree sizes ranged from 5 to 17, with an average of 9 members per family. Genotyping was done via massively parallel sequencing using the ForenSeq™ DNA Signature Prep kit (Verogen Inc., San Diego, CA, USA) and will be discussed in full elsewhere (M. Colucci, B. Rolf, N. A. Sheehan, M. A. Jobling, in preparation). Samples were collected with informed consent. For the purposes of the current study, we consider only the length-based genotypes from 27 autosomal STRs contained in Plex B of this kit. Allele frequencies are based on the European dataset in PopSTR (http://spsmart.cesga.es/popstr.php[34,35]) and downloaded from the Familias website (https://familias.no/download).

The performance analysis in Section 5.2, that leads to the blind search in the following section, is based on simulated data assuming the same set of loci as the real data, i.e., the 27 autosomal STR loci described above. This set of STR markers is also used in the simulations for the last example in the results section.

To demonstrate the use of X-chromosomal markers, data are simulated based on 12 X-chromosomal STR markers included in the kit "Investigator Argus X-12", with frequencies taken from an Argentinian database [36]. This is the most widely used kit for forensic applications.

4.2. Implementation

The analyses in this paper are all performed using R code developed by the authors. The code is available from the first author on request. The main engine of the code is an implementation of the parametric version of the likelihood function. This efficiently computes likelihoods for a series of relationships and converts to LRs and posterior probabilities. The code builds on the R libraries *pedtools*, *ribd*, *forrel* and *pedmut* developed by Magnus Dehli Vigeland, freely available from CRAN [37].

4.3. Assumptions

All the equations above are based on (1) which is only valid under certain assumptions. Firstly, the population is assumed to be in Hardy-

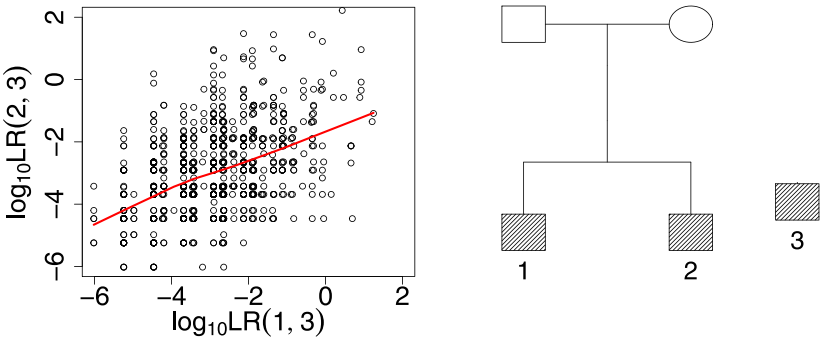

Fig. 6. Figure corresponding to the correlation discussion in Section 5.1. Left: Scatter plot of LRs of simulated data for two siblings and an unrelated individual. Red line shows regression line. Right: Pedigree used for simulation of data, identifying the id labels 1, 2 and 3 in left panel.

Weinberg Equilibrium (HWE) and in Linkage Equilibrium (LE). Secondly, mutations are ignored. Mutation rates are usually small, and the errors induced by ignoring them in likelihood calculations are typically negligible [38]. However, for a parent offspring (PO) relationship, i.e. $\kappa = (0, 1, 0)$, the likelihood will be zero if the two samples have genotypes at any locus that are incompatible with this hypothesis, e.g. $g_A = aa$ and $g_B = bb$. For this special case, there is a simple formulation of the likelihood that incorporates mutation (see [10]). This likelihood formula is applied throughout the paper when the likelihood of $\kappa = (0, 1, 0)$ is computed. An extended stepwise mutation model is implemented, with mutation rates of 10^{-3} and $5 \cdot 10^{-6}$ for mutation of integer and non-integer alleles, respectively and a mutation range of 0.1. For further details on this mutation model, see paper by Simonsson and Mostad [39]. We ignore allele drop-ins and drop-outs, null alleles and genotyping errors.

5. Results

The first example shows that the LRs in a blind search are not independent. The second example demonstrates how to evaluate the performance of a blind search such as we present in the third example. We then carry out a blind search on X-chromosomal markers before showing how inbreeding can be accommodated.

5.1. Correlation between LRs in a blind search

In this example, we show by simulation a case where the LRs of a blind search are correlated. Consider the pedigree in Fig. 6 and the hypotheses H_1 stating a sibling relationship and H_0 unrelated. Let $LR_{1,3}$ denote the likelihood ratio when individual 1 is compared to 3 and define $LR_{2,3}$ analogously. We use 10 independent loci, each with 10 alleles and equal allele frequencies of 0.1. Note that the LRs are random variables. We simulate 1000 sets of DNA profiles for the three shaded individuals of the pedigree in Fig. 6. The values of $LR_{1,3}$ and $LR_{2,3}$ are computed for each simulation. The results are shown in the scatter plot in Fig. 6, the red line denoting a regression line.

The estimated correlation between the logarithmic values of $LR_{1,3}$ and $LR_{2,3}$ is 0.484. This shows that the LRs are not independent. In other words, the outcome of different comparisons cannot be interpreted independently if one individual is involved in several comparisons. We elaborate on the implications of this correlation in the discussion.

5.2. From FWER to choice of LR threshold

In Section 5.3, we carry out a blind search among 65 individuals, genotyped for a set of 27 STR loci. Here, we present the preliminary evaluation required to obtain optimal LR thresholds for that search.

The first step is to decide on an acceptable value of α. From this value of α we can decide on an upper limit of the FPR and then the corresponding optimal LR threshold. For a blind search of $n = 65$ individuals, with the requirement that $\alpha \leq 0.05$, Equation (8) gives an upper limit for the false positive rate of $FPR_{0.05} = 2.404 \cdot 10^{-5}$.

The next step is to analyse how the FPR and TPR relate to each other for this particular set of markers. This depends on what hypotheses we test in the blind search. In the following example, we consider the hypotheses H_1: PO, H_2: S, H_3: H/U/G and H_4: FC, all against H_0: UN. We therefore consider these hypotheses here when estimating FPR and TPR.

Fig. 7 shows ROC curves for the different hypotheses. The values for $\widehat{FPR}$ and $\widehat{TPR}$ are estimated from simulated data, as described in Section 3.2. For H_1: PO, we only obtained estimates of FPR smaller than 10^{-7},

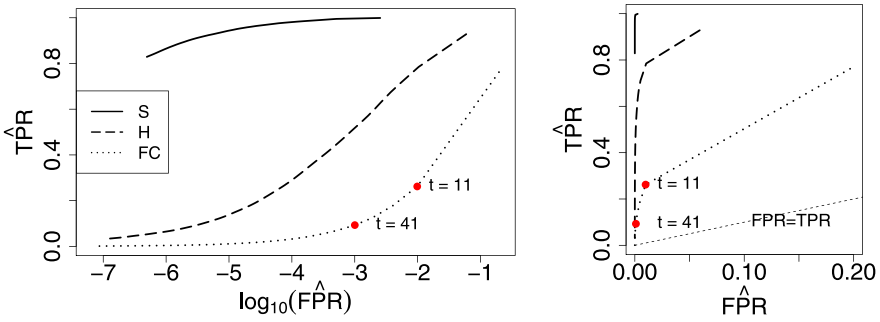

Fig. 7. ROC curves for the analysis performed in Section 5.2. The hypothesis H_1 stating S, H and FC and H_0 unrelated, using 27 STR markers. ROC curves from simulated data. A threshold of 11 corresponds to an estimated false positive rate of about 0.01 and an estimated true positive rate of about 0.26 for FC. The right figure shows the same estimated ROC curves and the line FPR = TPR, with an untransformed first axis.

Table 3
Optimal thresholds for different relationships, with corresponding $\widehat{\text{FPR}}$ and $\widehat{\text{TPR}}$, for the analysis performed in Section 5.2. For $\alpha = 0.05$ (left table) and $\alpha = 0.1$ (right table), for blind search with $n = 65$ individuals.

$\alpha = 0.05 \Rightarrow \text{FPR}_\alpha = 2.404 \cdot 10^{-5}$

	t	$\widehat{\text{FPR}}$	$\widehat{\text{TPR}}$
PO	65531	$5.439 \cdot 10^{-8}$	1.000
S	771	$2.296 \cdot 10^{-5}$	0.961
H	2501	$2.365 \cdot 10^{-5}$	0.184
FC	421	$1.689 \cdot 10^{-5}$	0.014

$\alpha = 0.1 \Rightarrow \text{FPR}_\alpha = 4.808 \cdot 10^{-5}$

	t	$\widehat{\text{FPR}}$	$\widehat{\text{TPR}}$
PO	65531	$5.439 \cdot 10^{-8}$	1.000
S	311	$4.713 \cdot 10^{-5}$	0.972
H	1551	$4.770 \cdot 10^{-5}$	0.232
FC	251	$4.421 \cdot 10^{-5}$	0.022

Table 4
Results of the blind search among $n = 65$ individuals in Section 5.3 with $\alpha = 0.05$ where N_1 denotes the total number of pairs in the sample with the tested relationship, TP is the number of these pairs with a LR above the threshold, and FP is the number of unrelated individuals with a LR above the threshold. The last row gives the number of other (differently related) pairs with an LR above the threshold.

	PO	S	H/U/G	FC
N_1	46	13	64	21
TP	43	12	12	0
FP	0	1	2	0
H_1 Claimed, other true	4	53	57	59

with a corresponding estimated TPR of 0.999 or higher. This shows that the LR comparing PO to UN is high when the true relationship is PO and low otherwise. Parent offspring and unrelated individuals are easily distinguished as expected and so we have omitted this curve from the graph.

The ROC curves show the estimated properties of a single computation of the LR, for the respective hypotheses, for this specific set of STR markers. The curves do not depend on the number of individuals in the blind search.

The last step in the performance analysis is to identify the optimal threshold, by minimising $\text{ER}(t)$, with the constraint $\widehat{\text{FPR}} \leq \text{FPR}_\alpha$. The highest optimal thresholds for $\alpha = 0.05$ and $\alpha = 0.1$ are listed in Table 3.

5.3. Blind search with real data

In this example we do a blind search on the data described in Section 4.1. The data set contains 65 DNA profiles. A blind search among these profiles results in 2080 pairwise comparisons. We want to test the hypotheses H_1: PO, $\kappa_1 = (0, 1, 0)$, H_2: S, $\kappa_2 = (0.25, 0.5, 0.25)$, H_3: H/U/G, $\kappa_3 = (0.5, 0.5, 0)$, H_4: FC, $\kappa_4 = (0.9375, 0.0625, 0)$, against H_0: UN, $\kappa_0 = (1, 0, 0)$. In the previous section, we obtained optimal thresholds for blind searches with these hypotheses (Table 3). A stepwise mutation model is implemented in the evaluation of PO.

Table 4 summarises the blind searches performed on the real data. This table is possible to construct because we know the true relationship for each pair from the pedigree information. In practice, only the sum of the last three rows (for each relationship) would be known.

For PO, we are left with a list of 47 hits. 43 of these are true PO, while 4 of the 47 hits are pairs of individuals with another relationship. 3 pairs with true PO relationship are not detected. By lowering the threshold, the remaining 3 pairs could have been detected. However, the probability of obtaining false positives increases by decreasing the threshold. For S, only one true sibling pair is not detected and there is only one false positive. However, the list of hits contains 66 pairs of individuals, 53 of these having another relationship.

We conclude that the summary in Table 4 is consistent with the performance evaluation shown in Table 3. PO can easily be distinguished from UN. The more distant the tested relationship, the lower the power to distinguish it from unrelated. With the obtained optimal thresholds, the number of false positives stays low as desired. For each hypothesis tested, the list of pairs warranting further investigation comprises those in the final row of Table 4, i.e., those who do not have the tested relationship and who are also not unrelated.

Table 5
LR values for seven pairs of the blind search in Section 5.3. Values for H, U and G are the same and shown in the column H/U/G. Values smaller than 10^{-6} are set to 0.

	PO	S	H/U/G	FC	UN	True
1	$\mathbf{5.181 \cdot 10^{10}}$	$\mathbf{1.205 \cdot 10^{8}}$	$\mathbf{1.825 \cdot 10^{7}}$	$\mathbf{5.593 \cdot 10^{4}}$	1	PO
2	353.460	$\mathbf{1.544 \cdot 10^{8}}$	$\mathbf{3.886 \cdot 10^{5}}$	$\mathbf{5.189 \cdot 10^{3}}$	1	S
3	0	0.681	57.572	20.519	1	H
4	0	$5.017 \cdot 10^{-3}$	4.156	4.0984	1	U
5	0	0.030	13.269	16.916	1	G
6	0	$1.115 \cdot 10^{-4}$	0.163	1.375	1	FC
7	0	$1.821 \cdot 10^{-6}$	0.022	0.349	1	UN

Table 6
Posterior probabilities, computed from the LR values of Table 5, when applying a flat prior, i.e., $\pi_i = 1/7$ for $i = 0, \ldots, 6$, as described in Section 5.4. Values for H, U and G are the same and shown in the column H/U/G. Probabilities smaller than 10^{-6} are set to 0.

	PO	S	H/U/G	FC	UN	True
1	**0.997**	0.002	0.0004	$1.076 \cdot 10^{-6}$	0	PO
2	$2.272 \cdot 10^{-6}$	**0.993**	0.002	$3.336 \cdot 10^{-5}$	0	S
3	0	0.0003	**0.295**	0.105	0.005	H
4	0	$2.86 \cdot 10^{-4}$	**0.237**	0.233	0.057	U
5	0	$5.15 \cdot 10^{-4}$	0.230	**0.293**	0.017	G
6	0	$3.90 \cdot 10^{-5}$	0.057	**0.480**	0.349	FC
7	0	$1.29 \cdot 10^{-6}$	0.0162	0.246	**0.706**	UN

5.4. Analysis of posterior probabilities

The result of each of the blind searches performed in Section 5.3 is a list of pairs with a LR above the threshold. Some pairs of individuals may appear in several of the lists, while other pairs may not be present in any of the lists. In this example, we turn to Bayesian analysis to further investigate specific pairs.

Table 5 shows LR values for 7 pairs from the above blind search. Values above the LR thresholds are in bold font. The rightmost column gives the true relationship. Only the first two pairs have LR values above the thresholds given in the left table of Table 3 corresponding to $\alpha = 0.05$. For pairs 3–7, the LRs are low, some below 1, indicating that a UN relationship is more plausible than the alternative hypothesis.

Next we calculate posterior probabilities to see if it is possible to infer a relationship for the different pairs. LR thresholds are not required for this. Table 6 shows posterior probabilities for the different hypotheses, with flat prior probabilities, i.e., $\pi_i = 1/7$ for $i = 0, \ldots, 6$. The highest probability for each pair is in bold and corresponds to the true relationship for several of the pairs. For example, the LRs in Table 5 comparing S, H/U/G and FC against UN for the second pair were all above the relevant LR thresholds. The posterior probability of S is close to 1, now making it possible to correctly infer this relationship. For pairs 3, 4 and 5, the highest posterior probabilities are just below 0.3. Even though the corresponding relationship is the most probable, a posterior probability of 0.3 is maybe not high enough to allow firm conclusions to be drawn.

The relationships H, U and G are indistinguishable in the parametric framework presented in Section 2. Also posterior probabilities with a flat

Table 7
Posterior probabilities with informative priors, as described in Section 5.4. Probabilities smaller than 10^{-6} are set to 0.

	PO	S	H	U	G	FC	UN	True
1	**0.999**	$6.57 \cdot 10^{-4}$	$3.06 \cdot 10^{-5}$	$2.52 \cdot 10^{-4}$	$2.07 \cdot 10^{-4}$	0	0	PO
2	$8 \cdot 10^{-6}$	**0.988**	$7.65 \cdot 10^{-4}$	0.006	0.005	$5.36 \cdot 10^{-5}$	0	S
3	0	0.001	0.038	**0.317**	0.259	0.072	0.312	H
4	0	$2.94 \cdot 10^{-5}$	0.007	0.062	0.051	0.039	**0.841**	U
5	0	$1.26 \cdot 10^{-4}$	0.017	0.143	0.117	0.116	**0.608**	G
6	0	0	$3.42 \cdot 10^{-4}$	0.003	0.002	0.015	**0.979**	FC
7	0	0	$4.78 \cdot 10^{-5}$	$3.95 \cdot 10^{-4}$	$3.23 \cdot 10^{-4}$	0.004	**0.995**	UN

prior as in Table 6 can not differentiate between them. Additional information, preferably objective, needs to be considered.

Suppose now that we have knowledge of how many pairs of the different relationships are present among the DNA profiles. This could be the case in a plane crash with a known passenger list. There are 1867 unrelated pairs, 46 parent-offspring pairs, 13 sibling pairs, 4 half sibling pairs, 33 avuncular pairs, 27 grandparental pairs and 21 first cousin pairs. The remaining 69 pairs have other more distant relationships not investigated here. The prior probabilities are then $\pi_0 = 0.928$ (UN), $\pi_1 = 0.023$ (PO), $\pi_2 = 0.006$ (S), $\pi_3 = 0.002$ (H), $\pi_4 = 0.016$ (U), $\pi_5 = 0.013$ (G) and $\pi_6 = 0.010$ (FC).

Posterior probabilities using these more informative priors are shown in Table 7. The prior probability of a PO relationships is $\pi_1 = 0.023$, i.e., there is a chance of 2.3% that a pair of individuals has a PO relationship. The corresponding posterior probability for the first pair is 0.999. The genetic data give such strong support to PO, that even though the prior probability is low, the posterior probability of this relationship is approximately 1.

In this blind search (as in most other blind searches), most pairs of individuals are unrelated, making the prior probability of UN close to 1 and the others low. This requires the LRs for the other relationships to be high in order to be supported by the posterior probabilities. For the relationships H/U/G and FC, the LR of the true relationships against UN is typically low. The combination of priors and LRs makes the posterior probability of UN high while the posterior probability of the true relationship remains low.

For this reason, this particular set of prior probabilities, even though objective, does not help us to distinguish between the H, U and G relationships in these data.

5.5. Blind search with X-chromosomal markers

Because a male has only one X-chromosome, paternal half sisters (HSP) must inherit the same X-chromosome from their common father. Their second X-chromosomes, inherited from their respective mothers, are not IBD (since their mothers are unrelated), and hence, the IBD coefficients for a HSP relationship are $\kappa = (0, 1, 0)$. The IBD coefficients for maternal half siblings (HSM), whether considering X-chromosomal or autosomal markers, are $\kappa = (0.5, 0.5, 0)$. In the following example, we show with simulated data how X-chromosomal markers can distinguish between HSP and HSM.

We simulated genotypes for 12 X-chromosomal STR markers, for the shaded individuals in Fig. 8. Genotypes are simulated for each locus independently, by gene-dropping through the pedigree structure. More specifically, genotypes are sampled for the founders of the pedigree (the parents) according to the allele population frequencies and passed down through the pedigree assuming the rules of Mendelian inheritance. Only the resulting genotypes of the offsprings are kept for the applications in this example. Table 8 presents the average posterior probabilities over

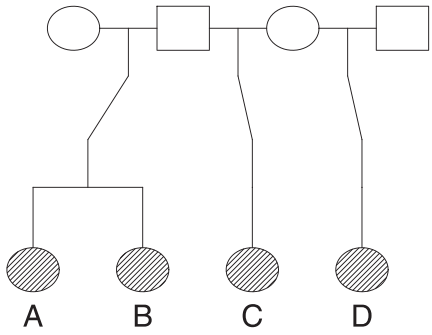

Fig. 8. Pedigree connecting the individuals of the analysis in Section 5.5. Marker data are simulated for the four daughters to demonstrate blind search with X-chromosomal markers.

100 simulations, for the relationships PO, S, HSP, HSM and UN, for the six possible comparisons between the individuals A, B, C and D. A flat prior $\pi_i = 1/5$ for $i = 0, \ldots, 4$ is assumed.

The evidence in favour of C-D being HSM, shown in bold in Table 8, could not be obtained using autosomal markers. Since we are using a flat prior, the LR comparing maternal to paternal half sibs can be found from the posterior probability ratio, $0.81327/0.01916 = 42.4$. This value may not be decisive on its own, but supplements other evidence. Note that HSP cannot be distinguished from PO using X-chromosomal markers alone as the row for the comparison A-C confirms. Age information, autosomal marker data or other non-DNA data may solve such cases.

5.6. Half siblings with inbred founder

Computations of LRs and posterior probabilities are restricted to a limited set of predefined pedigree relationships in many current software implementations. The parametric form of the LR given in (3) enables us to compute LRs and do blind search for any pairwise

Table 8
Posterior probabilities averaged over 100 simulations for the comparisons between the four daughters in Fig. 8.

	PO	S	HSP	HSM	UN
A-B	0.039	0.921	0.039	0.001	0.000
A-C	0.475	0.039	0.475	0.012	$2 \cdot 10^{-5}$
A-D	0.000	0.000	0.000	0.154	0.846
B-C	0.471	0.045	0.471	0.012	$2 \cdot 10^{-5}$
B-D	0.000	0.000	0.000	0.146	0.854
C-D	0.019	0.001	0.019	**0.813**	0.147

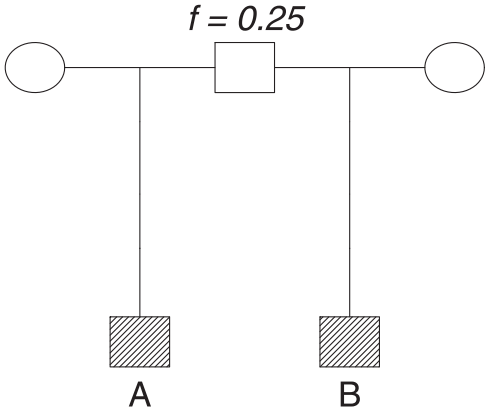

Fig. 9. Half sibling pedigree with founder inbreeding assumed in the analysis in Section 5.6.

relationship. In this example we show how background inbreeding can be modelled and how this can be taken into account in the Bayesian framework.

Assume a set of DNA profiles among which we want to do a blind search. The number of profiles is not important. The pedigrees connecting the individuals are unknown, but we know that the individuals come from a population where inbreeding is common. We consider the hypotheses H_1: PO, H_2: S, H_3: H and H_4: H with founder inbreeding $f = 0.25$, all against H_0: UN. The relationship in H_4 is shown in Fig. 9. Individuals A and B are outbred paternal half siblings, with the father being inbred with an inbreeding coefficient $f = 0.25$. This value of f corresponds to extreme inbreeding where the parents of the father are siblings. The IBD coefficients for the half sibling relationship are $\kappa = (0.375, 0.625, 0)$.

We consider one pair with true relationship H_4. A total of 100 simulations of DNA profiles for this pair is performed. LRs and posterior probabilities, with a flat prior $\pi_i = 1/5$, $i = 0, \ldots, 4$ are computed for each simulation. Mean values of the posterior probabilities for the hypotheses $H_0, \ldots, H_4$, are: $\bar{p}_1 = 0.017$, $\bar{p}_2 = 0.094$, $\bar{p}_3 = 0.374$, $\bar{p}_4 = 0.495$ and $\bar{p}_0 = 0.019$. It can be seen that the mean posterior probability of hypothesis H(f) is about 0.5, making it possible to distinguish it from the half sibling relationship without inbreeding.

The coefficient of inbreeding in this example is quite high. Lower values of f make the pair genetically more similar to half siblings without inbreeding, and distinguishing these relationships becomes harder without additional information. This high degree of inbreeding may be more relevant for non-human applications.

6. Discussion

The topic of this paper is blind search, a procedure used to search for pairwise relationships among a set of unidentified DNA profiles. Each pairwise comparison is similar to a kinship test performed, for instance, to resolve a paternity dispute. In the paper, we focus mainly on issues related to multiple testing. For this reason we will not discuss Hardy-Weinberg equilibrium and other assumptions that our applications share with other applications in forensic genetics. For instance, it is not obvious how evidence from different DNA sources like autosomal markers and X-chromosomal markers should be combined. However, this challenge is no different for a blind search than for a kinship test and is therefore not addressed here.

Case workers must decide on how the results of a blind search should be evaluated and reported. The context, or specific application, is obviously not irrelevant. In a DVI application, a false identification is likely to be a more serious error than missing an identification. To account for this, the metric for determining the threshold in (7) allows a weight to be specified which would penalise false identifications. Other applications, such as screening a database for relatives prior to estimating allele frequencies, may not require a weighting for errors. If costs can be specified for the possible errors, optimal decision rules can be derived as explained in Chapter 8.1 of [10]. However, there is hardly ever an objective way to balance the two errors that can occur and so specification of weights or costs may not be a viable option. We have used the unweighted form of the metric throughout. We have only considered binary decisions (corresponding to $t_0 = t_1$) as stated in the beginning of Section 2.4. One could drop this requirement and declare a test to be inconclusive if $t_0 < \mathrm{LR} < t_1$. In this case, a cost for making no decision would have to be added and (7) modified accordingly.

We only presented one method to determine an optimal threshold based on the distance illustrated in Fig. 4 although several alternatives are available [31]. Results using different approaches were practically identical for the examples we presented and so we chose not to discuss the thresholds based on the other metrics. Furthermore, other approaches to obtain ROC curves could be considered. For instance, there are several ways to smooth ROC curves. It is also possible to provide confidence bands for the ROC curves and study the impact of assumptions. This has been explored in previous papers [40]. Fig. 6 shows that the LRs from a blind search may be correlated when the same individual is involved in two comparisons. This has several implications. In particular, the results of different comparisons cannot be interpreted independently. Intuitively, we may get a high LR if unrelated individuals A and B happen to share a rare allele. Another individual C, who is a close relative of A, is likely to share this allele IBD with A and so we can also expect a high LR when comparing B and C. Importantly, the methods used to control the overall error rate must allow for dependence and for this reason we used the Bonferroni bound (8) as an upper limit for the FWER. Another frequently used measure to control the overall error rate in multiple testing scenarios is the false discovery rate (FDR) [14]. When controlling the FDR, the outcome of each test is based on p-values. However, conventional significance testing based on p-values are not recommended to evaluate the strength of DNA evidence in forensic genetics [41,42].

Furthermore, a blind search will not necessarily provide a globally consistent 'solution' in the sense that the LRs may support impossible combinations of relationships, like one individual having two mothers. An interesting extension to this paper would be to investigate alternative search strategies that may improve the results of a blind search. One strategy could be to do the search sequentially, where hypotheses to be tested depend on the outcome of the previous pairwise comparisons. For instance, if individuals A and B are classified as PO and A and C as PO, then it would be logical to test if B and C are siblings, half siblings or a grandparent-grandchild pair. There are also methods and software for pedigree reconstruction, see Chapter 8 of [37]. Finally, the true relationship may not be among the alternatives considered. This is also true for the Bayesian approach.

A Bayesian interpretation might seem more appropriate than the frequentist alternative for blind search applications than for a kinship case. The alternative, based on the LR, is designed to deal with only two hypotheses. If there are several hypotheses, a reference hypothesis must be specified. The posterior probabilities reported using a Bayesian approach make comparison of several competing hypotheses simpler as they are between 0 and 1. However, as always, a prior is needed for the Bayesian approach and the choice of prior may be crucial. If DNA is of poor quality, leading to few markers being typed, or if the competing hypotheses specify relationships that are very close to each other, conclusions may hinge on the choice of the prior.

An important aspect of this paper is the use of the parametric representation of relationships. This enables us to investigate any admissible pairwise relationship between two outbred individuals. By defining founder inbreeding in a pedigree structure, as shown in Fig. 3, background inbreeding can also be modelled [22]. Rather than proposing specific alternative relationships, we could simply estimate the coefficients describing the relationship. In the outbred case, these estimates can be plotted in the IBD triangle in Fig. 2 which would indicate where these relationships lie in relation to the well known relationships. For instance, pairs with estimates close to (0.25, 0.25) could be classified as siblings.

Throughout, we have restricted attention to pairwise testing. In principle, the blind search can be extended to search for relationship between triplets. However, the parametric approach based on the Jacquard coefficients then becomes impractical. The number of parameters needed to describe the relationship between three individuals increases, from 2 to 15 in the outbred case.

Issues to do with reporting DNA evidence are currently of key interest as evidenced by the so-called "DNA database controversy" (see [11] and references therein). The main message of this paper is that there are also problems related to multiple testing in kinship analyses which cannot be ignored.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank Burkhard Rolf for DNA samples. Margherita Colucci was supported by an iCASE MIBTP Biotechnology and Biological Sciences Research Council PhD studentship, grant ref. BB/M01116X/1, partnered by DNA Worldwide, and by a Short Term Fellowship of the International Society for Forensic Genetics (ISFG). We thank NUCLEUS Genomic Services at the University of Leicester for access to DNA sequencing. This research used the SPECTRE High Performance Computing Facility at the University of Leicester for data analysis.

Appendix A. Table of genotype probabilities

Table 9
The conditional probability $P(G|J_i)$ of a pair of genotypes $G = (g_A, g_B)$, given a Jacquard state J_i. The symbols a, b, c and d represent different alleles, with population frequencies p_a, p_b, p_c and p_d respectively.

(g_A, g_B)	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9
(aa, aa)	p_a	p_a^2	p_a^2	p_a^3	p_a^2	p_a^3	p_a^2	p_a^3	p_a^4
(aa, bb)	0	$p_a p_b$	0	$p_a p_b^2$	0	$p_a^2 p_b$	0	0	$p_a^2 p_b^2$
(aa, ab)	0	0	$p_a p_b$	$2p_a^2 p_b$	0	0	0	$p_a^2 p_b$	$2p_a^3 p_b$
(aa, bc)	0	0	0	$2p_a p_b p_c$	0	0	0	0	$2p_a^2 p_b p_c$
(ab, aa)	0	0	0	0	$p_a p_b$	$2p_a^2 p_b$	0	$p_a^2 p_b$	$2p_a^3 p_b$
(bc, aa)	0	0	0	0	0	$2p_a p_b p_c$	0	0	$2p_a^2 p_b p_c$
(ab, ab)	0	0	0	0	0	0	$2p_a p_b$	$p_a p_b (p_a + p_b)$	$4p_a^2 p_b^2$
(ab, ac)	0	0	0	0	0	0	0	$p_a p_b p_c$	$4p_a^2 p_b p_c$
(ab, cd)	0	0	0	0	0	0	0	0	$4p_a p_b p_c p_d$

Appendix B. Importance sampling

Importance sampling is a method that can be used to approximate small probabilities. We first introduce the indicator function,

$$I(\mathrm{LR} > t) = \left\{ \begin{array}{ll} 1, & \text{if } \mathrm{LR} \geq t, \\ 0, & \text{if } \mathrm{LR} < t. \end{array} \right\}$$

The expectation of I becomes

$$\begin{aligned} \mathrm{E}(I(\mathrm{LR} \geq t)) &= 0 \cdot \mathrm{P}(\mathrm{LR} < t) + 1 \cdot \mathrm{P}(\mathrm{LR} \geq t) \\ &= \mathrm{P}(\mathrm{LR} \geq t) \\ &= \mathrm{FPR} \end{aligned}$$

It is therefore valid to say that $\mathrm{FPR} = \mathrm{E}(I(\mathrm{LR} \geq t))$. Then consider the expression for the expected value in a more general sense. The value of the function I is dependent on the value of the LR, which is a function of the genotypes G of the DNA profiles. The probability distribution of G is governed by the relationships that has generated the data. For this consideration, we assume that this relationship is either H_0 or H_1. Denote by X the values that I can take on. We then have

$$\mathrm{E}(I(\mathrm{LR} \geq t)) = \sum_j X_j \cdot \mathrm{P}(G_j | \mathrm{H}_0) \approx \frac{1}{N} \sum_{i=1}^{N} I(\mathrm{LR}_i^{\mathrm{H}_0} \geq t).$$

In the last expression, the expected value is estimated by the sample mean of I, from a set of N simulations. The genotypes G, and then also X, are distributed according to H_0, which is indicated by the superscript of the LR. Then, consider the opposite probability distribution, $\mathrm{P}(G|\mathrm{H}_1)$, where the genotypes are distributed according to H_1. As long as $\mathrm{P}(G_j|\mathrm{H}_0) = 0$ whenever $\mathrm{P}(G_j|\mathrm{H}_1) = 0$, we can write

$$\mathrm{E}(I(\mathrm{LR} \geq t)) = \sum_j X_j \cdot \frac{\mathrm{P}(G_j | \mathrm{H}_0)}{\mathrm{P}(G_j | \mathrm{H}_1)} \mathrm{P}(G_j | \mathrm{H}_1) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{I(\mathrm{LR}_i^{\mathrm{H}_1} \geq t)}{LR_i^{\mathrm{H}_1}}.$$

Using this method, the LR is sampled under the wrong hypothesis (H_1), instead of the desired hypothesis (H_0). The bias this introduces is adjusted for by the weight $\mathrm{LR}^{\mathrm{H}_1}$. An estimate of FPR is then

$$\widehat{\mathrm{FPR}} = \frac{1}{N} \sum_{i=1}^{N} \frac{I(LR_i^{\mathrm{H}_1} \geq t)}{LR_i^{\mathrm{H}_1}}.$$

References

[1] B. Olaisen, M. Stenersen, M. M, Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster, Nat. Genet. 15 (1997) 402–405.
[2] B. Bertoglio, P. Grignani, P. Di Simone, N. Polizzi, D. De Angelis, C. Cattaneo, A. Iadicicco, P. Fattorini, S. Presciuttini, C. Previderè, Disaster victim identification by kinship analysis: the Lampedusa October 3rd, 2013 shipwreck, Forensic Sci. Int.: Genet. 44 (2020) 102–156.
[3] L. Olivieri, D. Mazzarelli, B. Bertoglio, D. De Angelis, C. Previderè, P. Grignani, A. Cappella, S. Presciuttini, C. Bertuglia, P. Di Simone, Challenges in the identification of dead migrants in the Mediterranean: the case study of the Lampedusa shipwreck of October 3rd 2013, Forensic Sci. Int. 285 (2018) 121–128.
[4] C.H. Brenner, Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities, Forensic Sci. Int. 157 (2006) 172–180.
[5] C.H. Brenner, B.S. Weir, Issues and strategies in the DNA identification of World Trade Center victims, Theor. Popul. Biol. 63 (2003) 173–178.
[6] T.J. Parsons, R.M.L. Huel, Z. Bajunović, A. Rizvić, Large scale DNA identification: the ICMP experience, Forensic Sci. Int.: Genet. 38 (2019) 236–244.
[7] S. Palomo-Díez, A. Esparza-Arroyo, M. Tirado-Vizcaíno, J. Velasco Vázquez, A. M. López-Parra, C. Gomes, C. Baeza-Richer, E. Arroyo-Pardo, Kinship analysis and allelic dropout: a forensic approach on an archaeological case, Ann. Hum. Biol. 45 (2018) 365–368.
[8] S. Palomo-Díez, A.M. López-Parra, C. Gomes, C. Baeza-Richer, A. Esparza-Arroyo, E. Arroyo-Pardo, Kinship analysis in mass graves: evaluation of the Blind Search tool of the Familias 3.0 Software in critical samples, Forensic Sci. Int.: Genet. Suppl. Ser. 5 (2015) e547–e550.
[9] T.J. Pemberton, C. Wang, J.Z. Li, N.A. Rosenberg, Inference of unexpected genetic relatedness among individuals in hapmap phase iii, Am. J. Hum. Genet. 87 (2010) 457–464.
[10] T. Egeland, D. Kling, P. Mostad, Relationship Inference with Familias and R: Statistical Methods in Forensic Genetics, Academic Press, 2015.
[11] G. Storvik, T. Egeland, The DNA database search controversy revisited: bridging the Bayesian–frequentist gap, Biometrics 63 (2007) 922–925.
[12] F.R. Bieber, C.H. Brenner, D. Lazer, Finding criminals through DNA of their relatives, Science (2006).
[13] M. Kruijver, R. Meester, K. Slooten, Optimal strategies for familial searching, Forensic Sci. Int.: Genet. 13 (2014) 90–103.
[14] J.D. Storey, The positive false discovery rate: a Bayesian interpretation and the q-value, Ann. Stat. 31 (2003) 2013–2035.
[15] D. Kling, A.O. Tillmar, T. Egeland, Familias 3–extensions and new functionality, Forensic Sci. Int.: Genet. 13 (2014) 121–127.
[16] T. Egeland, P.F. Mostad, B. Mevåg, M. Stenersen, Beyond traditional paternity and identification cases: selecting the most probable pedigree, Forensic Sci. Int. 110 (2000) 47–59.
[17] A. Jacquard, Genetic information given by a relative, Biometrics (1972) 1101–1114.
[18] N.A. Sheehan, T. Egeland, Structured incorporation of prior information in relationship identification problems, Ann. Hum. Genet. 71 (2007) 501–518.
[19] S. Wright, Coefficients of inbreeding and relationship, Am. Nat. 56 (1922) 330–338.
[20] E.A. Thomson, The estimation of pairwise relationships, Ann. Hum. Genet. 39 (1975) 173–188.
[21] E.A. Thompson, A restriction on the space of genetic relationships, Ann. Hum. Genet. 40 (1976) 201–204.
[22] M.D. Vigeland, T. Egeland, Handling founder inbreeding in forensic kinship analysis, Forensic Sci. Int.: Genet. Suppl. Ser. 7 (2019) 780–781.
[23] H.K. Brustad, T. Egeland, The impact of ignoring inbreeding in pairwise kinship evaluations, Forensic Sci. Int.: Genet. Suppl. Ser. 7 (2019) 462–464.
[24] M.D. Vigeland, Relatedness coefficients in pedigrees with inbred founders, J. Math. Biol. 81 (2020) 185–207.
[25] A.O. Tillmar, P. Mostad, Choosing supplementary markers in forensic casework, Forensic Sci. Int.: Genet. 13 (2014) 128–133.
[26] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (2006) 861–874.
[27] N. Pinto, L. Gusmão, A. Amorim, X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial, Forensic Sci. Int.: Genet. 5 (2011) 27–32.
[28] D. Kling, A. Tillmar, T. Egeland, P. Mostad, A general model for likelihood computations of genetic marker data accounting for linkage, linkage disequilibrium, and mutations, Int. J. Leg. Med. 129 (2015) 943–954.
[29] D. Kling, B. Dell'Amico, A.O. Tillmar, FamLinkX–implementation of a general model for likelihood computations for X-chromosomal marker data, Forensic Sci. Int.: Genet. 17 (2015) 1–7.
[30] M. Kruijver, Efficient computations with the likelihood ratio distribution, Forensic Sci. Int.: Genet. 14 (2015) 116–124.
[31] M. Rota, L. Antolini, Finding the optimal cut-point for Gaussian and Gamma distributed biomarkers, Comput. Stat. Data Anal. 69 (2014) 1–14.
[32] A.C. Tamhane, Y. Hochberg, C.W. Dunnett, Multiple test procedures for dose finding, Biometrics (1996) 21–37.
[33] M. Steffens, C. Lamina, T. Illig, T. Bettecken, R. Vogler, P. Entz, E. Suk, M.R. Toliat, N. Klopp, A. Caliebe, SNP-based analysis of genetic substructure in the German population, Hum. Hered. 62 (2006) 20–29.
[34] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magarinos, L. Porras, T. Tvedebrink, J. Amigo, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, A. Freire-Aradas, Analysis of global variability in 15 established and 5 new european standard set (ess) strs using the ceph human genome diversity panel, Forensic Sci. Int.: Genet. 5 (2011) 155–169.
[35] J. Amigo, C. Phillips, T. Salas, L.F. Formoso, A. Carracedo, M. Lareu, pop.str - an online population frequency browser for established and new forensic strs, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 361–362.
[36] M. García, C.I. Catanesi, G.A. Penacino, L. Gusmão, N. Pinto, X-chromosome data for 12 STRs: Towards an Argentinian database of forensic haplotype frequencies, Forensic Sci. Int.: Genet. 41 (2019) e8–e13.
[37] M.D. Vigeland, Pedigree Analysis in R, Academic Press, 2021.
[38] T. Egeland, N. Pinto, A. Amorim, Exact likelihood ratio calculations for pairwise cases, Forensic Sci. Int.: Genet. 29 (2017) 218–224.
[39] I. Simonsson, P. Mostad, Stationary mutation models, Forensic Sci. Int.: Genet. 23 (2016) 217–225.
[40] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinforma. 12 (2011) 1–8.
[41] M. Kruijver, R. Meester, K. Slooten, p-Values should not be used for evaluating the strength of DNA evidence, Forensic Sci. Int.: Genet. 16 (2015) 226–231.
[42] D.W. Gjertson, C.H. Brenner, M.P. Baur, A. Carracedo, F. Guidet, J.A. Luque, R. Lessig, W.R. Mayr, V.L. Pascali, M. Prinz, ISFG: recommendations on biostatistics in paternity testing, Forensic Sci. Int.: Genet. 1 (2007) 223–231.