



# Pairwise relatedness testing in the context of inbreeding: expectation and variance of the likelihood ratio

Hilde Kjelgaard Brustad<sup>1</sup> · Magnus Dehli Vigeland<sup>2</sup> · Thore Egeland<sup>1</sup>

Received: 13 January 2020 / Accepted: 10 September 2020 / Published online: 28 September 2020  
© The Author(s) 2020

## Abstract

In this paper we investigate various effects of inbreeding on the likelihood ratio (LR) in forensic kinship testing. The basic setup of such testing involves formulating two competing hypotheses, in the form of pedigrees, describing the relationship between the individuals. The likelihood of each hypothesis is computed given the available genetic data, and a conclusion is reached if the ratio of these exceeds some pre-determined threshold. An important aspect of this approach is that the hypotheses are usually not exhaustive: The *true* relationship may differ from both of the stated pedigrees. It is well known that this may introduce bias in the test results. Previous work has established formulas for the expected value and variance of the LR, given the two competing hypotheses and the true relationship. However, the proposed method only handles cases without inbreeding. In this paper we extend these results to all possible pairwise relationships. The key ingredient is formulating the hypotheses in terms of Jacquard coefficients instead of the more restricted Cotterman coefficients. While the latter describe the relatedness between outbred individuals, the more general Jacquard coefficients allow any level of inbreeding. Our approach also enables scrutiny of another frequently overlooked source of LR bias, namely background inbreeding. This ubiquitous phenomenon is usually ignored in forensic kinship computations, due to lack of adequate methods and software. By leveraging recent work on pedigrees with inbred founders, we show how background inbreeding can be modeled as a continuous variable, providing easy-to-interpret results in specific cases. For example, we show that if true siblings are subjected to a test for parent-offspring, moderate levels of background inbreeding are expected to inflate the LR by more than 50%.

**Keywords** Kinship analysis · Inbred founders · IBD triangle · Jacquard coefficients · Likelihood ratios

## Introduction

The conventional approach to forensic kinship testing includes formulating two hypotheses and calculating a likelihood ratio (LR) based on genetic data from genotyped individuals. Practice differs between countries and laboratories, but

typically the LR or some version of it is included when the case is reported. The conclusion based on the LR may be flawed when the true pedigree connecting the individuals of interest differs from the pedigrees considered by the hypotheses. As an example, consider a standard paternity case, where the prosecution asserts that a certain man is the father of a child, while the defense claims that the man and the child are unrelated. The truth, on the other hand, may be that the man is the child's uncle. A special case of incorrect hypotheses occurs when inbreeding is not accounted for. For example, if the alleged father is inbred, and this is ignored when formulating the hypotheses, this may significantly bias the LR. One aim of this paper is to investigate and quantify this effect.

Slooten and Egeland derived explicit equations for the expected value and variance of the LR [1]. They also extended this to cases where the true relationship differs from those stated in the hypotheses [2]. However, in both of these works only non-inbred individuals were considered.

---

✉ Hilde Kjelgaard Brustad  
hilde.brustad@nmbu.no

Magnus Dehli Vigeland  
m.d.vigeland@medisin.uio.no

Thore Egeland  
thore.egeland@nmbu.no

<sup>1</sup> Norwegian University of Life Sciences, 1432 Aas, Norway

<sup>2</sup> Department of Medical Genetics, University of Oslo, PB 4956 Nydalen, 0424 Oslo, Norway

An important contribution of this paper is the extension of these results to general pairwise relationships. In particular, we show that exact expressions for the expected value and variance of the LR can be obtained also in cases with inbreeding. The expressions are in general more involved than in the non-inbred case, and not as easy to interpret. However, we derive interesting and practical results in important special cases.

A parametric approach to modeling background inbreeding in kinship testing was recently introduced [3], employing the concept of inbred founders [4]. To exemplify, consider a pair of paternal half siblings, whose father is assigned an inbreeding coefficient  $f$ . As  $f$  increases from 0 to 1, the relationship between the half siblings becomes genetically indistinguishable from that between parent and child. We extend the theoretical framework of [1, 2] to pedigrees with inbred founders. As a result, the impact of background inbreeding on the expectation and variance of the LR can be studied based on exact expressions. In cases where the amount of inbreeding is unknown, we can still provide guidance on the expected values for the LR. Our approach conveniently allows a continuous range of possible true alternatives rather than a discrete set of specific alternatives. To arrive at explicit results of practical interest, we restrict attention to pairwise relationships. Furthermore, as in the work of Slooten and Egeland, we ignore mutations, dropouts, and silent alleles and we assume Hardy-Weinberg Equilibrium (HWE). However, we explain how deviation from HWE can be modeled by the so called theta ( $\theta$ ) correction.

R scripts and functions used to obtain numerical results in this paper are gathered in a R library (see the “R implementation” section). Pedigree likelihoods and marker simulations are performed with the forrel package [3].

This paper is organized in the following manner: After establishing some terminology and notation we review the main results of [2] regarding the expected value and variance of the LR for non-inbred pairs of individuals. We then proceed to extend these results to general pairwise relationships, including relationships in pedigrees with background inbreeding. Several worked examples follow, including a simulation study comparing our formulas with real-life results. Finally, we discuss some consequences of this work and how it relates to other aspects of forensic genetics.

## Definitions and notation

A central concept for measuring genetic relatedness is that of *identity by descent* (IBD). Two alleles are said to be IBD relative to a given pedigree if they are identical by state and

originate from the same ancestral allele within the pedigree [5].

## Coefficients of inbreeding and kinship

The *coefficient of inbreeding*  $f$ , introduced by Wright [6], is the probability that an individual is autozygous at a given autosomal locus, i.e., that the two homologous alleles are IBD. This is the same as the *kinship coefficient*  $\varphi$  between the parents of the same individual, defined as the probability that a random allele from the mother is IBD to a random allele from the father at the same locus.

Founders of a pedigree are conventionally assumed to be unrelated and non-inbred. Following [3] we relax the second assumption, allowing an arbitrary inbreeding coefficient  $f$  to be assigned to any founder individual. For a given pedigree with  $N$  founders, we denote the set of founder inbreeding coefficients by  $\mathbf{f} = (f_1, f_2, \dots, f_N)$ .

Background inbreeding in human populations is normally low, but may exceed 5% in certain cases [7, 8]. In forensic case work inbreeding is common, ranging from consanguineous marriages between cousins,  $f = 1/16$  or lower, to incestuous relationships between siblings or parent-child, both with  $f = 1/4$ . In breeding applications values closer to 1 may occur.

## Jacquard coefficients and likelihood of a pedigree

The kinship coefficient is a coarse measure of relatedness; for instance, it has the same value for a parent-child relationship as for full siblings. A more refined measure is given by the nine *Jacquard coefficients* [9]  $\mathbf{\Delta} = (\Delta_1, \dots, \Delta_9)$ , also called the *condensed identity coefficients*. These are the expected relative frequencies of the

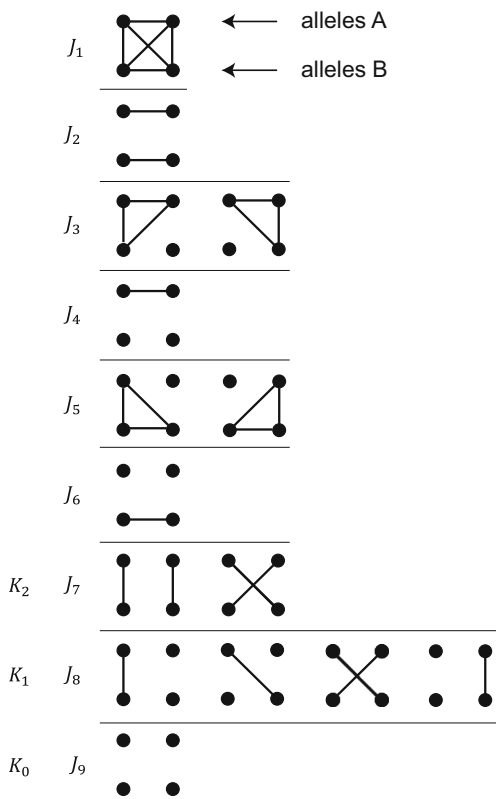
*Jacquard states*  $J_1, \dots, J_9$  are depicted in Fig. 1. Alleles within each individual are unordered, and hence, several IBD configurations can correspond to the same Jacquard state. Furthermore,  $\mathbf{\Delta}$  is related to  $\varphi$  through

$$\varphi = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8.$$

The likelihood of two individuals being related according to  $\mathbf{\Delta}$ , given their genotypes  $G = (g_1, g_2)$  at a marker may be expressed by conditioning on the Jacquard state:

$$L(\mathbf{\Delta} | G) = \sum_{i=1}^9 \Delta_i P(G | J_i). \quad (1)$$

The conditional probabilities  $P(G | J_i)$  are listed in Table 1. These probabilities are found by direct calculations; for instance,  $P((aa, aa) | J_1) = p_a$  since  $J_1$  dictates that all four alleles are IBD.



**Fig. 1** The Jacquad states  $J_1, \dots, J_9$  representing all possible IBD patterns among the four alleles of two individuals at an autosomal locus. Each row of dots represents the two alleles of an individual. Connected dots indicate IBD. The states  $J_9, J_8$ , and  $J_7$  do not involve inbreeding and are sometimes denoted  $K_0, K_1$ , and  $K_2$

**IBD coefficients and inbred founders**

For two non-inbred individuals, the first six Jacquad coefficients are zero, and  $\Delta_9, \Delta_8$ , and  $\Delta_7$  reduce to the IBD coefficients  $\kappa = (\kappa_0, \kappa_1, \kappa_2)$  introduced by Cotterman [10]. They give the probabilities that, at a given autosomal

locus, the individuals share zero-, one-, and two-allele IBD, respectively. Note that  $\kappa_0 + \kappa_1 + \kappa_2 = 1$ , so  $\kappa$  can be represented in a two-dimensional triangle with axes  $\kappa_0$  and  $\kappa_2$ . Thompson [11] showed that the IBD coefficients are restricted to  $\kappa_1^2 \geq 4\kappa_0\kappa_2$ . This gives rise to an inadmissible region for the parameters, in gray in Fig. 2.

Although the IBD coefficients are only defined for non-inbred individuals, other members of the pedigree can be inbred. For example, a pair of half siblings remain outbred even if their shared parent is inbred. However, this inbreeding will affect the relatedness coefficients. Table 2 lists the kinship and the IBD coefficients for some common relationships, as functions of the founder inbreeding. The effects are visualized in Fig. 2. In the half sibling example, the genetic relationship approaches that of parent-child, as the founder inbreeding increases towards 1. Similarly, the IBD coefficients of full siblings with inbred parents may fall anywhere in the lightly shaded region towards the point of monozygotic twins.

**Review of previous results**

We next review the main results of [2] relevant for our work. In particular we restate the explicit formulas for the expectation and variance of the LR in the case of non-inbred individuals.

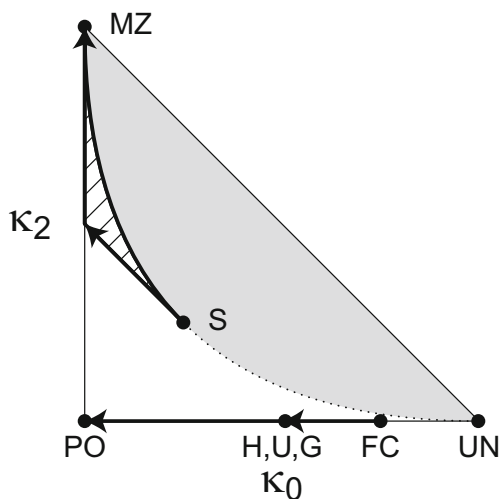
**The likelihood ratio as a random variable**

We consider a kinship test involving genetic data from two non-inbred individuals. Two hypotheses  $H_P$  and  $H_D$  about the relationship are to be compared using the LR. For our purposes, each hypothesis corresponds to a point in the IBD triangle, denoted by  $\kappa_P$  and  $\kappa_D$  respectively. However, the evidence may be generated from another pedigree, corresponding to a third point  $\kappa_T$ . We therefore have the

**Table 1** The conditional probability  $P(G | J_i)$  of a pair of genotypes  $G = (g_1, g_2)$ , given a Jacquad state  $J_i$

$G$	$J_1$	$J_2$	$J_3$	$J_4$	$J_5$	$J_6$	$J_7$	$J_8$	$J_9$
$(aa, aa)$	$p_a$	$p_a^2$	$p_a^2$	$p_a^3$	$p_a^2$	$p_a^3$	$p_a^2$	$p_a^3$	$p_a^4$
$(aa, bb)$	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	$4p_a p_b p_c p_d$

The symbols  $a, b, c$ , and  $d$  represent different alleles, with population frequencies  $p_a, p_b, p_c$ , and  $p_d$  respectively



**Fig. 2** The IBD triangle with location of some common relationships. The gray area is inadmissible. The arrows illustrate the effect of founder inbreeding in the cases given in Table 2. PO, parent-child; MZ, monozygotic twins; S, siblings; H, half siblings; U, avuncular; G, grandparent grandchild; FC, first cousins; UN, unrelated

following setup, comprising the competing hypotheses and the true relationship:

$$\begin{aligned}
 H_P : \quad \kappa &= \kappa_P = (\kappa_0^P, \kappa_1^P, \kappa_2^P) \\
 H_D : \quad \kappa &= \kappa_D = (\kappa_0^D, \kappa_1^D, \kappa_2^D) = (1, 0, 0) \\
 Truth : \quad \kappa &= \kappa_T = (\kappa_0^T, \kappa_1^T, \kappa_2^T).
 \end{aligned}$$

Reflecting standard practice, we will always use *unrelatedness* as the defense hypothesis, i.e.,  $\kappa_D = (1, 0, 0)$ . It should be noted, however, that this is not a theoretical requirement for the methods presented here.

The concept of the likelihood ratio as a random variable was discussed by Slooten and Egeland [1]. We review the basics here, presented in a slightly simpler notation sufficient for our purposes.

Denote by  $K_i, i = 0, 1, 2$ , the event that the individuals share exactly  $i$  alleles IBD. As shown in Fig. 1,  $K_0, K_1$ , and  $K_2$  are identical to the Jacquard states  $J_9, J_8$ , and  $J_7$  respectively. For fixed  $\kappa_P$  the likelihood ratio for a given pair of genotypes  $G = (g_1, g_2)$  can be written as

$$\begin{aligned}
 LR(G) &= \frac{P(G | H_P)}{P(G | H_D)} = \frac{P(G | \kappa_P)}{P(G | \kappa_D)} \\
 &= \sum_{i=0}^2 \kappa_i^P \frac{P(G | K_i)}{P(G | K_0)}. \tag{2}
 \end{aligned}$$

Note that the final transition was obtained by applying (1) in both the numerator and denominator. The probabilities  $P(G | K_i)$  are given in Table 1.

Now, viewing the genotypes as a random variable  $\mathcal{G}$ , we define the random variable  $\mathcal{LR} = LR(\mathcal{G})$ . Note that the distribution of  $\mathcal{G}$  is completely determined by  $\kappa_T$  (assuming HWE), hence the distribution of  $\mathcal{LR}$  is determined by  $\kappa_P$  and  $\kappa_T$ . If these parameters are clear from the context, we will suppress them in our notation; otherwise, we write

**Table 2** Relatedness coefficients as functions of founder inbreeding, in a selection of common relationships

Relationship	$\varphi$	$\varphi(f)$	$\kappa$	$\kappa(f)$
S	$\frac{1}{4}$	$\frac{1}{4}(1 + \frac{f_1+f_2}{2})$	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$	$\kappa_0(f_1, f_2) = \frac{1}{4}(1 - f_1)(1 - f_2)$ $\kappa_1(f_1, f_2) = \frac{1}{2}(1 - f_1)f_2$ $\kappa_2(f_1, f_2) = \frac{1}{4}(1 + f_1)(1 + f_2)$
H	$\frac{1}{8}$	$\frac{1}{8}(1 + f)$	$(\frac{1}{2}, \frac{1}{2}, 0)$	$\kappa_0(f) = \frac{1}{2}(1 - f)$ $\kappa_1(f) = \frac{1}{2}(1 + f)$ $\kappa_2(f) = 0$
U	$\frac{1}{8}$	$\frac{1}{8}(1 + \frac{f_1+f_2}{2})$	$(\frac{1}{2}, \frac{1}{2}, 0)$	$\kappa_0(f_1, f_2) = \frac{1}{2}(1 - \frac{f_1+f_2}{2})$ $\kappa_1(f_1, f_2) = \frac{1}{2}(1 + \frac{f_1+f_2}{2})$ $\kappa_2(f_1, f_2) = 0$
FC	$\frac{1}{16}$	$\frac{1}{16}(1 + \frac{f_1+f_2}{2})$	$(\frac{3}{4}, \frac{1}{4}, 0)$	$\kappa_0(f_1, f_2) = \frac{1}{4}(3 - \frac{f_1+f_2}{2})$ $\kappa_1(f_1, f_2) = \frac{1}{4}(1 + \frac{f_1+f_2}{2})$ $\kappa_2(f_1, f_2) = 0$

$\mathcal{LR}_{\kappa_P, \kappa_T}$ . In the special case when  $H_P$  equals the truth, i.e.,  $\kappa_P = \kappa_T$ , we may simplify  $\mathcal{LR}_{\kappa_P, \kappa_T}$  to  $\mathcal{LR}_{\kappa_P}$ .

Throughout, we assume the following condition to hold

$$P(G | H_P) > 0 \Rightarrow P(G | H_D) > 0. \tag{3}$$

In the present context, it means that all DNA profiles that can occur under  $H_P$ , can also occur under  $H_D$ . In our examples  $H_D$  specifies unrelated individuals, and then (3) holds. The condition also holds for mutation models provided all elements of the mutation matrix are positive. We do not model mutations in the work presented here, as practical exact expression are then no longer available. However, the implementation allows for general mutation models. Without (3), likelihood ratios could be infinite, i.e., not defined.

### Expected likelihood ratio

The expectation of  $\mathcal{LR}$  may be found by summing over all possible genotypes  $G$  in the standard way:

$$E(\mathcal{LR}) = \sum_G P(G)LR(G), \tag{4}$$

where  $P(G) = P(G | \kappa_T) = \sum_i \kappa_i^T P(G | K_i)$ . An exact expression for  $E(\mathcal{LR})$  when  $\kappa_P = \kappa_T$  was first derived in [1] and extended in [2] to apply when  $\kappa_P \neq \kappa_T$ . For the latter situation it was shown that, for a single marker with  $L$  alleles,

$$E(\mathcal{LR}) = \kappa_P \cdot A_0 \cdot (\kappa_T)^t, \tag{5}$$

where  $t$  denotes the vector transpose, and

$$A_0 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \frac{L+3}{4} & \frac{L+1}{2} \\ 1 & \frac{L+1}{2} & \frac{L(L+1)}{2} \end{pmatrix}. \tag{6}$$

Importantly, the expected value depends only on the number of alleles, not on the allele frequencies. Furthermore, the expectation is symmetric in  $\kappa_P$  and  $\kappa_T$ , so that

$$E(\mathcal{LR}_{\kappa_P, \kappa_T}) = E(\mathcal{LR}_{\kappa_T, \kappa_P}). \tag{7}$$

### Variance of the likelihood ratio

To derive the variance of  $\mathcal{LR}$  we apply the general formula  $\text{var}(\mathcal{X}) = E(\mathcal{X}^2) - E(\mathcal{X})^2$ . Since the last term follows from

Eq. 5, all that remains is to find the first term. Some notation is needed:

$$\begin{aligned} s_1 &= \frac{1}{16} \sum_{a < b} \left( \frac{p_a}{p_b} + \frac{p_b}{p_a} \right), \\ s_2 &= \sum_{a < b} \frac{1}{2p_a p_b}, \\ s_3 &= \sum_a \frac{1}{p_a}, \\ s_4 &= \frac{1}{4} \sum_{a < b} \left( \frac{1}{p_b} + \frac{1}{p_a} \right), \\ s_5 &= \sum_a \frac{1}{p_a^2}. \end{aligned}$$

Furthermore, supplementing the matrix  $A_0$  given in Eq. 6, we define matrices  $A_1$  and  $A_2$  by

$$A_1 = \begin{pmatrix} 1 & \frac{L+3}{4} & \frac{L+1}{2} \\ \frac{L+3}{4} & \frac{5L+3}{8} + s_1 & \frac{L(L+7)}{8} + 2s_1 \\ \frac{L+1}{2} & \frac{L(L+7)}{8} + 2s_1 & s_3 + s_4 \end{pmatrix} \tag{8}$$

$$A_2 = \begin{pmatrix} 1 & \frac{L+1}{2} & \frac{L(L+1)}{2} \\ \frac{L+1}{2} & \frac{L(L+7)}{8} + 2s_1 & s_3 + s_4 \\ \frac{L(L+1)}{2} & s_3 + s_4 & s_2 + s_5 \end{pmatrix} \tag{9}$$

It was shown in [2] that

$$E(\mathcal{LR}^2) = \sum_{i=0}^2 \kappa_i^P \kappa_P A_i (\kappa_T)^t;$$

hence, the complete variance expression becomes

$$\begin{aligned} \text{var}(\mathcal{LR}) &= \\ & \sum_{i=0}^2 \kappa_i^P \kappa_P A_i (\kappa_T)^t - (\kappa_P A_0 (\kappa_T)^t)^2. \end{aligned} \tag{10}$$

Contrary to the expected LR, the variance of the LR depends on the allele frequencies.

### Example: paternity testing

This example serves as an illustration of the above described expected LR and the corresponding hypotheses. Consider a paternity case, where a man is claimed to be the father of a child ( $H_P$ ). The truth is that a brother of the alleged father is the true father of the child. The hypotheses and the true relatedness are in terms of the IBD coefficients given as

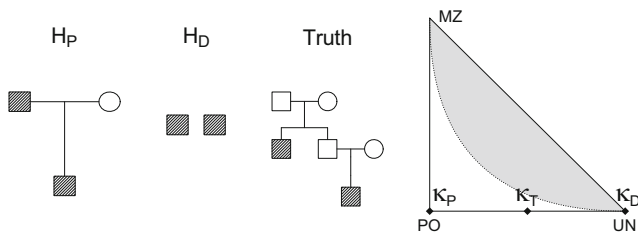
$$\begin{aligned} H_P : \quad \kappa &= \kappa_P &= (0, 1, 0) \\ H_D : \quad \kappa &= \kappa_D &= (1, 0, 0) \\ Truth : \quad \kappa &= \kappa_T &= (\frac{1}{2}, \frac{1}{2}, 0). \end{aligned} \tag{11}$$

Figure 3 illustrates the hypotheses in terms of pedigrees, and as points in the IBD triangle. Equation (5), with IBD coefficients as in Eq. 11, simplifies to

$$E(\mathcal{LR}) = \frac{L+7}{8}. \tag{12}$$

The variance of  $\mathcal{LR}$  becomes

$$\text{var}(\mathcal{LR}) = \frac{7L+9}{16} + \frac{s_1}{2} - \left(\frac{L+7}{8}\right)^2.$$



**Fig. 3** Pedigrees and location of IBD coefficients  $\kappa_P$ ,  $\kappa_D$ , and  $\kappa_T$  for a paternity case when the true relationship is avuncular

In the special case  $L = 2$ , and allele frequencies  $q$  and  $1 - q$ , the variance expression reduces to

$$\text{var}(\mathcal{LR}) = \frac{11}{64} + \frac{1}{32} \frac{(1 - q)^2 + q^2}{q(1 - q)}.$$

This expression is minimal when  $q = 0.5$  and becomes infinitely large when  $q$  or  $1 - q$  approaches 0. If no assumption is made for  $L$ , but all alleles are assumed equally frequent, the variance reduces to

$$\text{var}(\mathcal{LR}) = \frac{L(L + 12)}{64} - \frac{13}{64}. \tag{13}$$

Table 3 exemplifies these formulas for various realistic values of  $L$ , and compares the results with the corresponding values if  $H_P$  was true.

### Likelihood ratio for general pairwise relationships

In this section we extend the results reviewed above to relationships between any pairs of individuals. In particular we now allow inbreeding. For this to work we must pass from the IBD coefficients to the full set of Jacquard coefficients. For details regarding derivations of the results (see the Appendix).

### Expected likelihood ratio

We use the same setup for kinship testing as introduced previously, but in order to allow general inbreeding,

we now formulate our hypotheses using Jacquard coefficients,

$$\begin{aligned} H_P : \mathbf{\Delta} = \mathbf{\Delta}_P &= (\Delta_1^P, \dots, \Delta_9^P) \\ H_D : \mathbf{\Delta} = \mathbf{\Delta}_D &= (\Delta_1^D, \dots, \Delta_9^D) = (0, \dots, 0, 1) \\ Truth : \mathbf{\Delta} = \mathbf{\Delta}_T &= (\Delta_1^T, \dots, \Delta_9^T). \end{aligned}$$

Note that the defense hypothesis still corresponds to unrelatedness. We are interested in the likelihood ratio comparing  $H_P$  with  $H_D$  when the genotypes are generated by a pedigree with the Jacquard coefficients  $\mathbf{\Delta}_T$ . Equation (1) implies that

$$\begin{aligned} LR(G) &= \frac{P(G | \mathbf{\Delta}_P)}{P(G | \mathbf{\Delta}_D)} \\ &= \sum_{i=1}^9 \Delta_i^P \frac{P(G | J_i)}{P(G | J_9)}. \end{aligned} \tag{14}$$

As shown in the Appendix, the expected LR is

$$E(\mathcal{LR}_{\mathbf{\Delta}_P, \mathbf{\Delta}_T}) = \mathbf{\Delta}_P B_9 (\mathbf{\Delta}_T)^t, \tag{15}$$

where  $B_9$  is the symmetric  $9 \times 9$  matrix given in Table 4, whose elements are  $E(\mathcal{LR}_{J_i, J_j})$ , for  $1 \leq i, j \leq 9$ . As opposed to the non-inbred case, we see that the expected value in general depends on the allele frequencies.

### Variance of the likelihood ratio

In the Appendix matrices  $B_1, \dots, B_9$  are defined and it is shown that

$$E(\mathcal{LR}^2) = \sum_{i=1}^9 \Delta_i^P \mathbf{\Delta}_P B_i (\mathbf{\Delta}_T)^t. \tag{16}$$

From this we obtain the variance formula

$$\begin{aligned} \text{var}(\mathcal{LR}) &= \\ & \sum_{i=1}^9 \Delta_i^P \mathbf{\Delta}_P B_i (\mathbf{\Delta}_T)^t - (\mathbf{\Delta}_P B_9 (\mathbf{\Delta}_T)^t)^2. \end{aligned} \tag{17}$$

### Pairwise relationships with inbred founders

As previously explained, a set of inbreeding coefficients  $f$  can be assigned the founders of a pedigree to model background inbreeding. The Jacquard coefficients of any pair of pedigree members are then functions of  $f$ . It follows that the formulas for expectation and variance of  $\mathcal{LR}$

**Table 3** Expectation and variance of  $\mathcal{LR}$  in the paternity example of Fig. 3, for loci with 2, 10, and 50 alleles

Truth	$\kappa_P$	$\kappa_T$	E[LR]	$L = 2$	$L = 10$	$L = 50$
PO	(0, 1, 0)	(0, 1, 0)	$\frac{L+3}{4}$	1.250 (0.188)	3.250 (1.686)	13.250 (9.188)
U	(0, 1, 0)	( $\frac{1}{2}, \frac{1}{2}, 0$ )	$\frac{L+7}{8}$	1.125 (0.234)	2.125 (3.234)	7.125 (48.230)

The variances are computed assuming uniform allele frequencies. The bottom row ( $U$ ) shows the values when the true pedigree is uncle-nephew, as analyzed in the main text. For comparison, the top row shows the corresponding numbers when  $H_P$  is true

**Table 4** Elements of the symmetric matrix  $B_9$ , given as  $E(\mathcal{LR}_{J_i, J_j})$

	$J_1$	$J_2$	$J_3$	$J_4$	$J_5$	$J_6$	$J_7$	$J_8$	$J_9$
$J_1$	$\sum \frac{1}{p_a^2}$	$\sum \frac{1}{p_a}$	$\sum \frac{1}{p_a}$	$L$	$\sum \frac{1}{p_a}$	$L$	$\sum \frac{1}{p_a}$	$L$	1
$J_2$		$L^2$	$L$	$L$	$L$	$L$	$L$	1	1
$J_3$			$\frac{1}{2}(L + \sum \frac{1}{p_a})$	$L$	$L$	1	$L$	$\frac{L+1}{2}$	1
$J_4$				$L$	1	1	1	1	1
$J_5$					$\frac{1}{2}(L + \sum \frac{1}{p_a})$	$L$	$L$	$\frac{L+1}{2}$	1
$J_6$						$L$	1	1	1
$J_7$							$\frac{L(L+1)}{2}$	$\frac{L+1}{2}$	1
$J_8$								$\frac{L+4}{3}$	1
$J_9$									1

Each row represents  $J_i$ , a Jacquard state assumed by  $H_P$ , while each column presents  $J_j$ , the true Jacquard state

involving such pedigrees remain as in Eqs. 15 and 17, except that the parameters  $\Delta_P$  and  $\Delta_T$  must be updated.

Specifically, let  $f_P$  be a vector of founder inbreeding coefficients in the pedigree assumed by  $H_P$ , and  $f_T$  similarly in the true pedigree. The expectation and variance of  $\mathcal{LR}$  in this situation are then given by

$$E(\mathcal{LR}_{\Delta_P(f_P), \Delta_T(f_T)}) = \Delta_P(f_P) B_9 (\Delta_T(f_T))^t$$

and

$$\begin{aligned} & \text{var}(\mathcal{LR}_{\Delta_P(f_P), \Delta_T(f_T)}) \\ &= \sum_{i=1}^9 \Delta_i^P(f_P) \Delta_P(f_P) B_i (\Delta_T(f_T))^t \\ & \quad - (\Delta_P(f_P) B_9 (\Delta_T(f_T))^t)^2. \end{aligned}$$

Note that the matrices  $B_i$  only depend on  $L$  and the allele frequencies, and therefore are unchanged by founder inbreeding.

*Remark 1* It should be emphasized that the formulas (15) and (17) are needed only when at least one of the tested individuals are inbred in some of the involved pedigrees. If both are non-inbred, the simpler expressions (5) and (10) using IBD coefficients suffice. Importantly, this remains true if *other* members of the pedigree are inbred, as long as this does not lead to inbreeding in the tested individuals. In particular, founder inbreeding may be accounted for in Eqs. 5 and 10 simply by replacing  $\kappa_P$  and  $\kappa_T$  by  $\kappa_P(f_P)$  and  $\kappa_T(f_T)$  respectively.

**Founder inbreeding and  $\theta$  correction**

The conventional approach to background relatedness in forensics is the so called  $\theta$  correction [12]. In an inbred population, the composition of genotypes do not follow the Hardy-Weinberg principle, implying that the frequencies given in Table 1 no longer hold. The following approach

compensates for this by adjusting the allele frequencies. Without loss of generality we can assume that alleles observed are sampled sequentially. The probability that allele  $i$  is sampled as the  $j$ th allele is given by the *sampling formula*

$$p'_i = \frac{b_j \theta + \bar{\theta} p_i}{1 + (j - 2)\theta}, \tag{18}$$

where  $\bar{\theta} = 1 - \theta$  and  $b_j$  denotes the number of alleles of type  $i$  among the  $j - 1$  previously sampled. Note that for pairwise cases, the likelihood can be written

$$L(\Delta(f) | G, \theta) = \sum_{i=1}^9 \Delta_i(f) P(G | J_i, \theta), \tag{19}$$

where  $P(G | J_i, \theta)$  is calculated using Eq. 18. The matrices  $B_1, \dots, B_9$  then change with  $\theta$ , modifying the expectation and variance of the LR. This emphasises a fundamental difference between founder inbreeding and  $\theta$  correction:  $f$  modifies the relationship itself, while  $\theta$  only impacts the genotype probabilities.

**Example:  $\theta$  correction and founder inbreeding in a paternity case**

This example compares  $\theta$  correction to founder inbreeding. Consider first the hypothesis  $H_D$ : A and B are unrelated. Assume both individuals are homozygous a/a. Equation (18) gives the likelihood

$$L_\theta(H_D) = p_a(\theta + \bar{\theta} p_a) \frac{2\theta + \bar{\theta} p_a}{1 + \theta} \frac{3\theta + \bar{\theta} p_a}{1 + 2\theta}.$$

If rather than using  $\theta$  correction, we assign an inbreeding coefficient  $f$  to A, the likelihood becomes

$$L_f(H_D) = (f p_a + (1 - f) p_a^2) p_a^2.$$

Consider next the hypothesis  $H_{P1}$ : A is the father of B. Equation (18) now gives

$$L_{\theta}(H_{P1}) = p_a(\theta + \bar{\theta}p_a) \frac{2\theta + \bar{\theta}p_a}{1 + \theta}$$

and so the LR with  $\theta$  correction is

$$LR_{\theta} = \frac{L_{\theta}(H_{P1})}{L_{\theta}(H_D)} = \frac{1 + 2\theta}{3\theta + \bar{\theta}p_a}$$

The inbreeding coefficient approach gives

$$L_f(H_{P1}) = (fp_a + (1 - f)p_a^2)p_a$$

and  $LR_f = 1/p_a$ . Note that the LR does not depend on  $f$  and that this is true for all genotype combinations for A and B. The LRs for other genotype combinations for A and B with  $\theta$  correction are given in Table 10.8 in [13].

To illustrate (19) consider the hypothesis  $H_{P2}$ : A and B are paternal half siblings whose father is inbred. Table 2 then gives  $\Delta_8 = \frac{1}{2}(1 + f)$  and  $\Delta_9 = \frac{1}{2}(1 - f)$ , and by Eqs. 18 and 19 we may write down the likelihood for any genotype combinations. For instance, when A is homozygous  $a/a$  and B homozygous  $b/b$  the likelihood is

$$L(f, \theta) = \frac{1}{2}(1 - f)p_a(\theta + \bar{\theta}p_a) \frac{\bar{\theta}p_b}{1 + \theta} \frac{\theta + \bar{\theta}p_b}{1 + 2\theta}$$

The LR comparing  $H_{P2}$  with A and B being unrelated becomes  $\frac{1}{2}(1 - f)$ . If A and B share alleles, the LR will depend also on  $\theta$ .

### R implementation

Utilities to perform the computations in this paper are provided in a R library named InbredLR, available from the first author, building on several packages in the *ped suite*, notably *pedprobr* and *forrel* [3]. The core of InbredLR are functions that compute the expectation and variance of the likelihood ratio for pairwise relationships. The user can specify the parameters ( $\kappa$ ,  $f$  or  $\Delta$ ) or specify the pedigrees, possibly with inbred founders. A function for simulating marker data to estimate the distribution of LR is also provided, as well as a function for visualizing pedigrees  $H_P$  and  $H_D$  and the true pedigree and location of the corresponding IBD coefficients in the IBD triangle.

### Results

#### Paternity case for siblings with inbred founders

Consider two individuals who claim to be related as parent and offspring. Their true relationship is siblings and their parents coefficients of inbreeding are  $f_T = (f_1, f_2)$ . Figure 4 shows the case. This example can be relevant

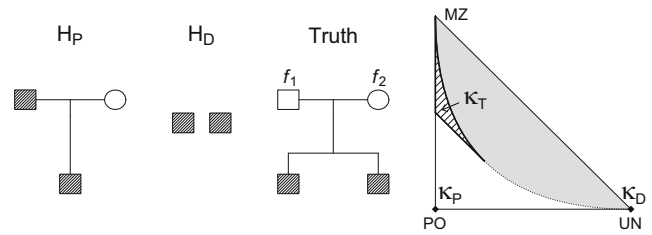


Fig. 4 Hypotheses involved in “Paternity case for siblings with inbred founders” and the location of the corresponding IBD coefficients  $\kappa_P$ ,  $\kappa_D$ , and  $\kappa_T$  in the IBD triangle

for family reunion cases, where a parent-child relationship would give right to residence permit, whereas a sibling relationship would not. In [14] such a case is considered.  $H_P$  and  $H_D$  and their true relationship are in terms of the IBD coefficients given as

$$\begin{aligned} H_P : \quad \kappa &= \kappa_P = (0, 1, 0) \\ H_D : \quad \kappa &= \kappa_D = (1, 0, 0) \\ Truth : \quad \kappa &= \kappa_T(f_T), \end{aligned} \tag{20}$$

where  $\kappa_T(f_T) = \kappa_T(f_1, f_2)$  are as in the first row of Table 2. Keeping in mind Remark 1, we apply (5) to find the expected LR:

$$E(\mathcal{LR}) = \frac{L - 1}{8}(f_1 + f_2) + \frac{L + 3}{4} \tag{21}$$

Figure 5 plots  $E(\mathcal{LR})$  as a function of the inbreeding level (assuming  $f_1 = f_2$ ), for a single locus with  $L = 2, 10$  and  $50$  alleles.

Without founder inbreeding,  $E(\mathcal{LR}) = (L + 3)/4$ . Interestingly, this is the same as the expectation if  $H_P$  was true, i.e., if the two individuals were in fact father and son (see first row of Table 3). The variance of  $\mathcal{LR}$  differs between the two cases, however (not shown here).

As the background inbreeding of the true sibling pedigree increases,  $E(\mathcal{LR})$  increases. The expected LR of the

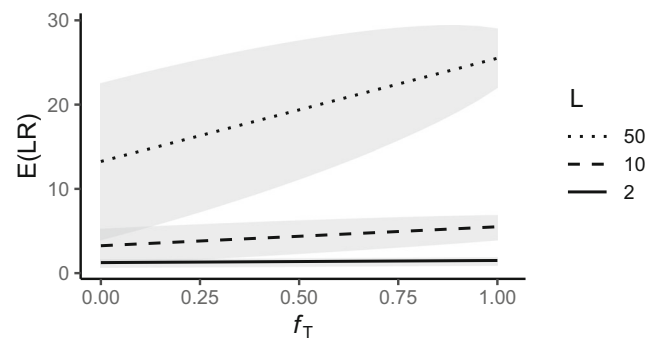


Fig. 5  $E(\mathcal{LR})$  as function of background inbreeding level  $f_T$  (assuming  $f_1 = f_2$ ), for  $L = 2, 10$ , and  $50$  alleles, for the paternity case in Fig. 4. The shaded area shows one standard deviation below and above  $E(\mathcal{LR})$ , for uniform allele frequencies



paternity case (and hence the trust in  $H_P$ ) is therefore higher if the true relatedness is siblings with background inbreeding, rather than the tested parent-child relationship. The variance of  $\mathcal{LR}$  decreases moderately for increasing founder inbreeding. For increasing number of alleles  $L$ , the slope of the expected LR increases.

The following calculation gives a simple approximation of the inflation in the expected LR caused by background inbreeding. Suppose  $f_1 = f_2 = f$ , and write (21) as  $\mu_0 + \mu_f$ , where  $\mu_0 = \frac{1}{4}(L + 3)$  is the expected LR without founder inbreeding, and  $\mu_f = \frac{1}{4}(L - 1)f$  is the expected contribution caused by founder inbreeding. Note that  $\mu_0 + \mu_f = (1 + \frac{\mu_f}{\mu_0})\mu_0$ , and that for  $L \geq 5$  we have  $\frac{\mu_f}{\mu_0} = \frac{L-1}{L+3}f \geq \frac{1}{2}f$ . This implies that with  $N$  independent markers, the total LR has expectation

$$[(1 + \frac{\mu_f}{\mu_0})\mu_0]^N \geq (1 + \frac{1}{2}f)^N \mu_0^N \geq (1 + \frac{1}{2}fN)\mu_0^N.$$

This means that a background inbreeding level  $f$  will inflate the expected LR by at least  $\frac{1}{2}fN$ . For example, if  $N = 20$  and  $f = 0.05$ , the inflation rate is greater than 50%.

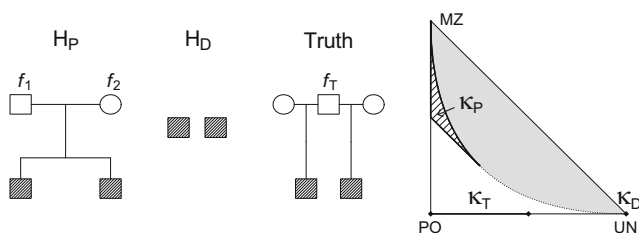
### Siblings and half siblings with founder inbreeding

Distinguishing between siblings and half siblings can be difficult based on unlinked markers. Mayor and Balding address the problem in [15], with focus on the number of loci needed. If the shared parent of the half siblings has inbreeding coefficient  $f_T > 0$ , the problem becomes even more interesting.

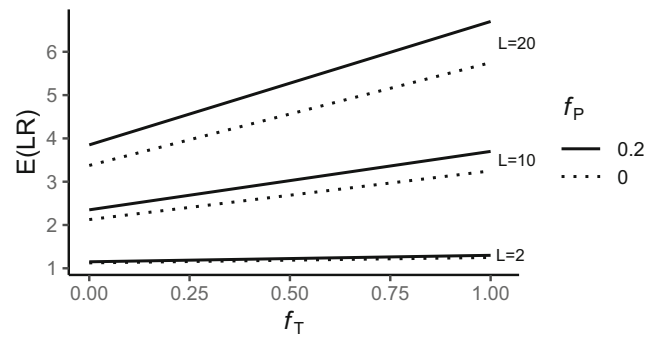
Consider the situation shown in Fig. 6. The hypotheses are

$$\begin{aligned} H_P : \quad \kappa &= \kappa_P(f_P) \\ H_D : \quad \kappa &= \kappa_D = (1, 0, 0) \\ Truth : \quad \kappa &= \kappa_T(f_T), \end{aligned} \tag{22}$$

where  $f_P = (f_1, f_2)$  are the parental inbreeding coefficients in the  $H_P$  pedigree and  $\kappa_P(f_P)$  and  $\kappa_T(f_T)$  are as in the first and second rows of Table 2, respectively.



**Fig. 6** The hypotheses involved in “Siblings and half siblings with founder inbreeding” and the location of the corresponding IBD coefficients  $\kappa_P$ ,  $\kappa_D$ , and  $\kappa_T$  in the IBD triangle



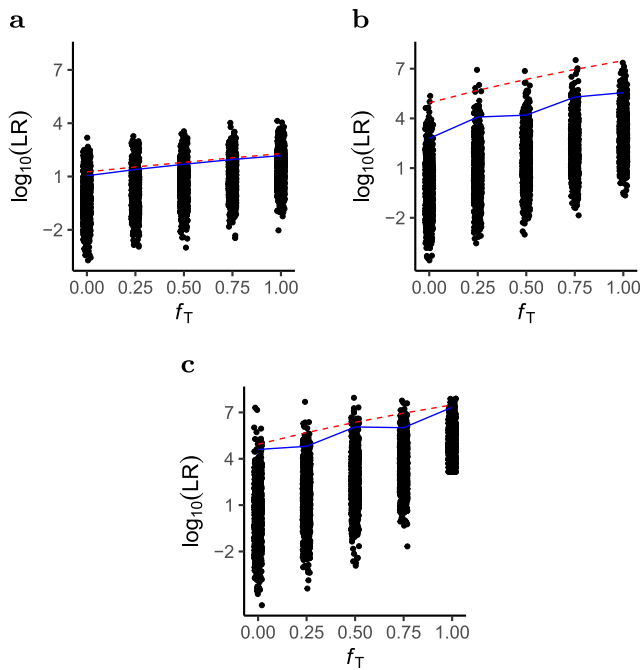
**Fig. 7**  $E(\mathcal{LR})$  for the case in Fig. 6 as functions of background inbreeding level  $f_T$ , for  $f_P = 0$  (dashed line) and  $f_P = 0.2$  (solid line), and  $L = 2, 10$ , and  $20$

This setup facilitates for modeling background inbreeding in both the true pedigree and in  $H_P$ . Equation (5) gives

$$E(\mathcal{LR}) = \frac{L - 1}{8} \left( \frac{(f_1 + f_2)(f_T + 1)}{2} + f_T \right) + \frac{L + 7}{8}. \tag{23}$$

In Fig. 7, the expectation of  $\mathcal{LR}$  is shown as a function of founder inbreeding  $f_T$  of the true half sibling pedigree, for  $H_P$  stating sibling pedigree with founder inbreeding  $f_P = 0$  and  $0.2$  (assuming  $f_1 = f_2$ ), and  $L = 2, 10$  and  $20$  alleles at a locus. For increasing values of  $f_T$ ,  $E(\mathcal{LR})$  increases, for all values of  $f_P$ , and the evidence in favor of a sibling relationship becomes stronger.

Consider next the situation when  $f_1 = f_2 = 0$ .  $H_P$  then assumes a sibling relationship without inbred founders. Figure 8 shows  $E(\mathcal{LR})$  (dashed line) and  $LR$  computations from 1000 sets of simulated data, as a function of  $f_T$ . The solid line gives the mean value of the simulated  $LR$ . The expected LR increases slightly as founder inbreeding increases. For Fig. 8a this seems to fit well with the mean values of the  $LR$ s from simulated data. These simulation assumes 13 loci, each of 3 alleles with allele frequencies 0.4, 0.3 and 0.3. In Fig. 8b, on the other hand, there is a substantial difference between  $E(\mathcal{LR})$  and the mean of the simulated  $LR$ s. These simulations use 13 CODIS markers with allele frequencies ranging from 0.0003 to 0.5378 (allele frequencies are available as a part of the R library InbredLR, see the “R implementation” section). Alleles with low frequencies will more seldom be present in the simulations. The expected LR only depends on the number of alleles at a locus, but because of the rare alleles, the simulations give in practice a lower number of alleles at these loci. The simulations in Fig. 8c use the same markers, but with uniform allele frequencies for alleles at a locus. The expectation of the LR is independent of the allele frequencies and is therefore not changed, but now the mean of the simulated  $LR$ s is closer to the expected value. Even though  $E(\mathcal{LR})$  is independent of the allele frequencies, the



**Fig. 8** Simulations of  $LR$  for the case in Fig. 6. Each figure shows 1000  $LR$  values, for five values of  $f_T$ , each calculated from a simulation of a complete set of genotypes for 13 loci. Solid lines show mean of simulated  $LR$ . Dashed lines show  $E(\mathcal{LR})$ . **a** Loci with 3 alleles with frequencies 0.4, 0.3 and 0.3. **b** CODIS loci with realistic allele frequencies. **c** CODIS loci with uniform allele frequencies

variance is not, and small allele frequencies increase the variance.

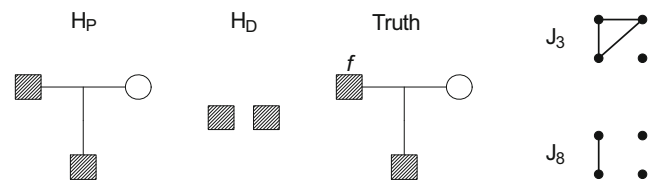
Finally, we offer an approximation of the inflation in the expected  $LR$  due to background inbreeding. For simplicity, we assume  $f_1 = f_2 = 0$  so that  $H_P$  states a normal sibling relationship. From Eq. 23 the expected  $LR$  is  $\mu_0 = \frac{1}{8}(L+7)$  if  $f_T = 0$ . On the other hand, if  $f_T > 0$ , the expected contribution to the  $LR$  is  $\mu_f = \frac{1}{8}(L-1)f_T$ . For  $L \geq 5$  we have  $\frac{\mu_f}{\mu_0} \geq \frac{1}{3}f_T$ , and it follows that

$$(\mu_0 + \mu_f)^N = \left[ \left(1 + \frac{\mu_f}{\mu_0}\right) \mu_0 \right]^N \geq \left(1 + \frac{1}{3}f_T\right) \mu_0^N.$$

A background inbreeding level of  $f_T$  will inflate the expected  $LR$  by at least  $\frac{1}{3}f_T N$ . For example, with  $N = 20$  and  $f_T = 0.05$ , the inflation rate is greater than 33%.

**Paternity case with inbreeding**

Consider a paternity case with hypotheses as shown in Fig. 9. The alleged father is indeed the true father and has inbreeding coefficient  $f$ . We will analyze the consequences



**Fig. 9** The hypotheses in a paternity case with inbreeding. To the far right are the Jacquard states with nonzero probability in the true relationship

of ignoring the inbreeding in  $H_P$ . The hypotheses are parameterized in the following way:

$$\begin{aligned} H_P : \quad \Delta &= \Delta_P = (0, 0, 0, 0, 0, 0, 0, 1, 0) \\ H_D : \quad \Delta &= \Delta_D = (0, 0, 0, 0, 0, 0, 0, 0, 1) \\ Truth : \quad \Delta &= \Delta_T(f_T) \\ &= (0, 0, f_T, 0, 0, 0, 0, 1 - f_T, 0). \end{aligned}$$

The expression for the expected  $LR$  simplifies considerably since most elements of  $\Delta_P$  and  $\Delta_T(f_T)$  are zero. Equation (15) gives

$$E(\mathcal{LR}) = \frac{L+1}{2}f_T + \frac{L+3}{4}(1-f_T),$$

and we see that  $E(\mathcal{LR})$  increases linearly from  $(L+3)/4$  to  $(L+1)/2$  as  $f_T$  goes from 0 to 1.

Consider next the variance. For brevity, we define

$$h(i, j, k) = E(\mathcal{LR}_{J_i, J_k} \cdot \mathcal{LR}_{J_j, J_k}). \tag{24}$$

Note that  $h(i, j, k)$  is invariant under permutations of  $i, j, k$ . Equation (16) gives

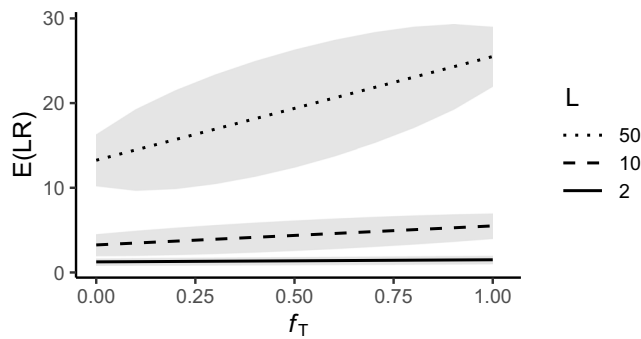
$$\begin{aligned} E(\mathcal{LR}^2) &= \Delta_3^T h(8, 8, 3) + \Delta_8^T h(8, 8, 8) \\ &= f_T h(8, 8, 3) + (1 - f_T) h(8, 8, 8). \end{aligned}$$

Slooten and Egeland [1] derived the term not involving inbreeding, i.e.,

$$h(8, 8, 8) = \frac{5L+3}{8} + \frac{s_3-L}{16}.$$

To derive the remaining term we condition on the zygosity of the son. If he is homozygous  $a/a$ , the father must also be  $a/a$  (recall that we are conditioning on Jacquard state  $J_3$ ). Conversely, if the son is heterozygous  $a/b$ , the father is equally likely to be  $a/a$  or  $b/b$ . This gives

$$\begin{aligned} h(8, 8, 3) &= \sum_a p_a^2 \frac{1}{p_a} \frac{1}{p_a} \\ &+ \sum_{a < b} 2p_a p_b \left( \frac{1}{2} \left(\frac{1}{2p_a}\right)^2 + \frac{1}{2} \left(\frac{1}{2p_b}\right)^2 \right) \\ &= L + \frac{1}{4} \sum_{a \neq b} \frac{p_b}{p_a^2} = \frac{3L + s_3}{4}. \end{aligned}$$



**Fig. 10**  $E(\mathcal{LR})$  as a function of  $f_T$  in the paternity case in Fig. 9, for a single marker with  $L = 2, 10$  and  $50$  alleles. Shaded area shows one standard deviation below and above  $E(\mathcal{LR})$ , for uniform allele frequencies

In summary,

$$\begin{aligned} \text{var}(\mathcal{LR}) = & \frac{3L + s_3}{4} f_T + \left( \frac{5L + 3}{8} + \frac{s_3 - L}{16} \right) (1 - f_T) \\ & - \left( \frac{L + 1}{2} f_T + \frac{L + 3}{4} (1 - f_T) \right)^2. \end{aligned} \tag{25}$$

This is a concave function with respect to  $f_T$ . Figure 10 shows  $E(\mathcal{LR})$  and one standard deviation on each side as a function of founder inbreeding  $f_T$ , for different number of alleles at a locus.

### Discussion

In testing theory, the formulation of hypotheses is crucial. Kinship problems, as considered in this paper, are no exception. The convention of kinship testing is to compare two specific relationships using the LR. In most applications other than kinship problems, the hypotheses together span many, if not all, alternatives. For instance, a common example is testing of HWE against *all* possible deviations from HWE. In forensic genetics,  $H_P$ : “paternity” is typically tested only against  $H_D$ : “unrelated,” not all other alternatives. For this reason, it becomes essential to study what happens when the truth is neither of these hypotheses.

A pairwise non-inbred relationship can be presented by a point in the IBD triangle (see Fig. 2), or in general by the Jacquard coefficients (see Fig. 1). We have presented two ways of expressing the hypotheses and the true relationship; (i) through the Jacquard coefficients, and (ii) background relatedness or founder inbreeding. These approaches let us investigate the LR for a continuous range of relationships and values of background relatedness. In both cases, the impact on the LR has been studied by deriving exact

expressions for its mean and variance. In the latter case, the required formula follows rather directly by extending results in [1] and [2]. Explicit formulas for the expected LR has been derived for several sets of relationships. In the case of Jacquard coefficients, the explicit formulas are complicated to derive, and they depend on allele frequencies. An exact expression is given also for the variance. However, as the variance depends on allele frequencies, simple closed formulas can only be derived in special cases. For general applications we rely instead on the exact numerical implementation freely available in the R library InbredLR accompanying this paper.

Equipped with the results of this paper, we can address the following question when presented with a standard LR comparing two completely specified hypotheses  $H_P$  and  $H_D$ : What if the true relationship between the individuals is not as stated by  $H_P$ ? Or this slightly different question: What if the true relationship is restricted to some particular region of the IBD triangle. Obviously, the LR can be re-evaluated to reflect the new specifications. However, the exact expressions for expectation and variance of the LR can in some cases directly allow for statements valid for a continuous range of alternatives. For instance, regions obtained by varying founder inbreeding have been displayed in Fig. 2. Assume a LR has been reported in a paternity case and that inbreeding in the father has been ignored. It is then useful to know that accounting for inbreeding would imply increase in the expected LR. This finding could be essential as there may not be data available to estimate the inbreeding coefficient for the father. Hence, exact LR calculation is not feasible.

Because the definition of “common ancestor” sometimes differs, there is a slight difference in the definition of IBD in the literature. The paper [16] gives three definitions of IBD: ancient IBD, recent IBD, and familial IBD. Our definition of IBD goes in the category of familial IBD, where “common ancestor” is restricted to a given pedigree.

The conventional approach to background relatedness in forensics is the so called theta ( $\theta$ ) correction [12]. Typical values are  $\theta \in (0.01, 0.03)$ . The  $\theta$  parameter applies on a population level. The genotype probabilities of all founders in the pedigree are modified compared with what HWE would give. Our approach does not model relatedness between founders, but offers a richer model of inbreeding, since individual inbreeding coefficients can be specified for each founder.

Several authors (see, e.g., [2] and the references therein) have discussed reporting the logarithm of the LR rather than the LR. Nice expressions like the ones presented for the expectation and the variance are then no longer available. In most cases, the LR is reported on the original scale. In some circumstances, as for paternity cases, the LR may be 0, and

then, the logarithm is not defined. Many papers including [17] study the distribution of  $\mathcal{Z} = \log(\mathcal{LR})$  by simulation. Equipped with the exact expressions of this paper,  $\mathcal{Z}$  could be analyzed without resorting to simulation, since the mean and variance of  $\mathcal{Z}$  can be derived from the counterparts for the LR. However, if some allele frequencies are close to 0,  $\mathcal{Z}$  is not well approximated by a normal distribution for a realistic number of markers. The reason for this is the large variance when allele frequencies are small. For instance, (25) shows an example where the expression for the variance include terms of the form  $1/p_a$  and these become large whenever the allele frequency  $p_a$  is small. A similar problem related to small allele frequencies is discussed in the result section. This demonstrates that the center of the  $\log(\mathcal{LR})$  distribution, calculated from the expectation of  $\mathcal{LR}$ , can be inaccurate. However, this criticism applies to the use of  $\mathcal{LR}$  instead of  $\log(\mathcal{LR})$  in general, and not specifically to the expectations. We maintain that results like the ones presented for the expectation and variance have considerable theoretical interest, but should be used with caution in practice.

This paper has mainly addressed the likelihood ratio and its properties. The exclusion probability (EP), the probability that genotypes will be incompatible with a claimed relationship, is also an important statistic. The impact of founder inbreeding on EP is discussed in [3].

Figure 4 illustrates a case where the true inbred relationship is not known, and Fig. 5 shows the corresponding expected LR for a single marker. Increasing the number of markers will, in this paternity case, increase the inflation of the expected LR. This means that adding more markers to the LR computation will not solve the problem. In general, with a sufficient number of markers, the Jacquard, IBD, or inbreeding coefficients can be estimated accurately, and the true relationship detected. If such additional marker data is not available, the impact of inbreeding can be studied as exemplified by a paternity case with unknown inbreeding earlier in the discussion and as illustrated in, e.g., Fig. 5. As addressed in the “Introduction” section, different scenarios can be investigated and LR results can be evaluated in light of the analyses of these scenarios.

The present paper does not consider linked markers. For independent loci, the inbreeding coefficients contain sufficient information to compute the Jacquard coefficients needed in our formulas for LR. While a similar approach is conceivable also for linked markers, this would involve multi-locus coefficients, which is outside the scope of this work.

**Funding** Open Access funding provided by Norwegian University of Life Sciences.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** None required as no data from humans are used.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix: Expectation and variance of LR

Below we derive the expressions for the expectation and variance of  $\mathcal{LR}$  in the general pairwise case. Let  $J_i$  denote Jacquard state  $i$  and  $\Delta_i^P$  and  $\Delta_i^T$  the probabilities of  $J_i$  according to the relationship stated by  $H_P$  and the true relationship respectively.  $\mathcal{LR}_{\Delta_P, \Delta_T}$  is then defined as the likelihood ratio comparing  $H_P: \Delta_P$  with  $H_D: \Delta_D$  when the marker data comes from the relationship  $\Delta_T$ . Similarly,  $\mathcal{LR}_{J_i, J_j}$  denotes the likelihood ratio comparing Jacquard state  $J_i$  with unrelated, i.e.,  $J_9$  when the marker data are generated by  $J_j$ .

Equation (15) follows by combining (1), (14), and (4)

$$\begin{aligned} E(\mathcal{LR}) &= \sum_G \left( \sum_{j=1}^9 \Delta_j^T P(G | J_j) \sum_{i=1}^9 \Delta_j^P \frac{P(G|J_i)}{P(G|J_9)} \right) \\ &= \sum_{i=1}^9 \sum_{j=1}^9 \Delta_i^P \Delta_j^T \left( \sum_G \frac{P(G|J_i)}{P(G|J_9)} P(G | J_j) \right) \quad (26) \\ &= \sum_{i=1}^9 \sum_{j=1}^9 \Delta_i^P \Delta_j^T E(\mathcal{LR}_{J_i, J_j}) \\ &= \Delta_P B_9 \Delta_T^t. \end{aligned}$$

In the case of no inbreeding, i.e.,  $\Delta_1 = \dots = \Delta_6 = 0$ , the above expression reduces to (5). The part of the  $9 \times 9$  matrix  $B_9$  corresponding to  $(J_7, J_8, J_9)$  coincides with the matrix given in Eq. 6. Since  $E(\mathcal{LR}_{J_i, J_j}) = E(\mathcal{LR}_{J_j, J_i})$ ,  $B_9$  is symmetric. The elements of  $B_9$  are found by direct calculation. For instance, entry (1, 1) equals

$$E(\mathcal{LR}_{J_1, J_1}) = \sum_a \frac{p_a}{p_a^4} p_a = \sum_a \frac{1}{p_a^2}.$$

Since the expectation has been calculated, to derive the variance it remains only to find

$$\begin{aligned}
 & E(\mathcal{LR}^2) \\
 &= \sum_G \left( \sum_{k=1}^9 \Delta_k^T P(G | J_k) \sum_{i=1}^9 \Delta_i^P \frac{P(G | J_i)}{P(G | J_9)} \right. \\
 &\quad \left. \sum_{j=1}^9 \Delta_j^P \frac{P(G | J_j)}{P(G | J_9)} \right) \\
 &= \sum_{i=1}^9 \sum_{j=1}^9 \sum_{k=1}^9 \Delta_i^P \Delta_j^P \Delta_k^T E(\mathcal{LR}_{J_j, J_i} \mathcal{LR}_{J_k, J_i}) \\
 &= \sum_{i=1}^9 \Delta_i^P \Delta_P B_i \Delta_i^T.
 \end{aligned}$$

The matrices  $B_1, \dots, B_9$  are symmetric  $9 \times 9$  matrices. The simplest of these matrices is  $B_9$ , given in Table 4. In general,  $B_i$  consists of the elements  $\{E(\mathcal{LR}_{J_j, J_i} \mathcal{LR}_{J_k, J_i})\}_{j,k=1, \dots, 9}$ . The values for  $i, j, k = 7, 8, 9$  have been provided in the “Review of previous results” section. Entry  $(j, k)$  of  $B_i$  is

$$\sum_G \frac{P(G | J_j) P(G | J_k)}{P(G | J_9) P(G | J_9)} P(G | J_i). \tag{27}$$

All matrices can in principle be found from the above expression, but exact calculations by hand become unpractical and exact numerical calculation is more reasonable.

**References**

1. Slooten KJ, Egeland T (2014) Exclusion probabilities and likelihood ratios with applications to kinship problems. *Int J Legal Med* 128(3):415–425
2. Egeland T, Slooten KJ (2016) The likelihood ratio as a random variable for linked markers in kinship analysis. *Int J Legal Med* 130(6):1445–1456

3. Vigeland MD, Egeland T (2019) Handling founder inbreeding in forensic kinship analysis. *Forensic Science International: Genetics Supplement Series*. <https://doi.org/10.1016/j.fsigs.2019.10.175>
4. Vigeland MD (2020) Relatedness coefficients in pedigrees with inbred founders. *J Math Biol* 81:185–207. <https://doi.org/10.1007/s00285-020-01505-x>
5. Thompson EA (2000) Statistical inference from genetic data on pedigrees. *IMS*
6. Wright S (1922) Coefficients of inbreeding and relationship. *The American Naturalist* 56:330–338
7. Buckleton J, Curran J, Goudet J, Taylor D, Thiery A, Weir BS (2016) Population-specific FST values for forensic STR markers: a worldwide survey. *Forensic Science International: Genetics* 23:91–100. <https://doi.org/10.1016/j.fsigen.2016.03.004>
8. Pemberton TJ, Rosenberg NA (2014) Population-genetic influences on genomic estimates of the inbreeding coefficient: a global perspective. *Human Heredity* 77(1-4):37–48. <https://doi.org/10.1159/000362878>
9. Jacquard A (1972) Genetic information given by a relative. *Biometrics* 28(4):1101–1114
10. Cotterman CW (1940) A calculus for statistico-genetics. Dissertation, The Ohio State University
11. Thompson EA (1976) A restriction on the space of genetic relationships. *Ann Hum Genet* 40(2):201–204
12. Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* 96(1–2):3–12
13. Buckleton J, Triggs CM, Walsh SJ (2005) Forensic DNA evidence interpretation. CRC Press, Florida
14. Gorlin JB, Polesky HF (2000) The use and abuse of the full-sibling and half-sibling indices. *Transfusion* 40(9):1148–1149
15. Mayor LR, Balding DJ (2006) Discrimination of half-siblings when maternal genotypes are known. *Forensic Sci Int* 159(2–3):141–147
16. Browning BL, Browning SR (2011) A fast, powerful method for detecting identity by descent. *The American Journal of Human Genetics* 88(2):173–182
17. Nothnagel M, Schmidtke J, Krawczak M (2010) Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. *Int J Legal Med* 124(3):205–215

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.