Available online at www.sciencedirect.com
SCIENCE DIRECT
Forensic Science International 159 (2006) 141–147
Forensic Science International
www.elsevier.com/locate/forsciint

Discrimination of half-siblings when maternal genotypes are known

Lianne R. Mayor, David J. Balding *

Department of Epidemiology and Public Health, Imperial College, St. Mary's Campus, Norfolk Place, London W2 1PG, UK

Received 8 July 2005; accepted 18 July 2005
Available online 8 September 2005

Abstract

Given the DNA profiles of two individuals and one parent (say the mother) of each, we present likelihood ratios (LRs) comparing the hypothesis that they have the same father with the hypothesis of unrelated fathers. If the individuals have the same mother, the problem is to distinguish full- from half-siblings, otherwise we are comparing a half-sibling relationship with unrelated. We simulate STR profiles at up to 60 loci, based on allele proportions observed at 15 loci in three populations, and use them to approximate misclassification rates both for binary classification (e.g. ''half-sib'' versus ''unrelated''), and when a third ''cannot say'' category is included. We find that reliable inferences in the absence of the mothers' profiles require many more STR loci than the 10–25 loci that are currently routinely available. However, profiling the two mothers conveys more discriminatory power than profiling the same number of additional loci in the individuals themselves. Our likelihood ratio formulas include a θ (or F_{st}) adjustment to allow for the individuals concerned to have recent shared ancestry (coancestry), relative to the population from which the allele frequency database is drawn. We illustrate that using an appropriate value of θ can reduce the average misclassification rate.

Keywords: Likelihood ratios; DNA profiling; STRs; Coancestry; Relatedness; Population structure; F_{st}

1. Introduction

Likelihood ratios (LRs) for two individuals being half-siblings rather than unrelated, based on their multi-locus DNA profiles, have previously been reported [1], but not when the DNA profiles of one parent of each individual are also available. Typically in practice, it is the mother that is known and the question of interest is whether the individuals have a common father. We assume this setting, but the same results apply if the fathers are known and the query concerns a possible common mother. Further, although we focus on individuals having different mothers (i.e. distinguishing half-siblings from unrelateds), the same-mother situation (i.e. distinguishing full- from half-sibs) is the same as when the mothers are different but happen to have the same genotype (except that the θ-adjustment changes).

A large number of STR loci is typically required to distinguish half-siblings from unrelated pairs of individuals, and we will see that including the mothers' STR profiles can greatly reduce the number of distinct loci that need to be used for genotyping. In fact, we present below simulation results indicating that the reduction in the number of loci required for the same discriminatory power is more than 50%, so that there is a (small) reduction in the total genotyping requirement when mothers are genotyped. We also consider the effect of a θ (=F_{st}) adjustment for coancestry, and find that an appropriate θ-adjustment can reduce the average of the two misclassification rates.

* Corresponding author. Tel.: +44 20 7594 3309; fax: +44 20 7402 2150.
E-mail address: d.balding@imperial.ac.uk (D.J. Balding).

0379-0738/$ – see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2005.07.007

2. Materials and methods

2.1. Single-locus likelihood ratios

Let H_1 denote the hypothesis that two individuals with unrelated mothers are half-siblings, and let H_0 denote that they are unrelated. Thus, under H_1, the individuals have the same father and under H_0 the two fathers are unrelated; we do not consider here the possibility that the individuals have different but related fathers. Under H_1 at a given locus, let $Z = 1$ if the two individuals share an allele identical by descent (ibd) from their father, otherwise $Z = 0$. Under Mendelian inheritance, $P(Z = 1) = P(Z = 0) = 1/2$. If $Z = 0$, then the genotypes are equally likely under H_0 and H_1, and so the relevant terms in the single-locus LR cancel to leave:

$$\mathrm{LR} = \frac{1}{2}\left(1 + \frac{P(\text{observed genotypes}|H_1, Z = 1)}{P(\text{observed genotypes}|H_0)}\right). \quad (1)$$

The half-sibling relationship is genetically the same as uncle–aunt, niece–nephew and grandparent–grandchild, and hence the LR for comparing any of these relationships with unrelated has the same form as (1), except that Z now refers to shared inheritance from the most recent common ancestor of the two individuals (in the case of grandparent–grandchild, this is the grandparent him/herself).

Ayres [1] has reported explicit forms of (1) when the mothers' profiles are unavailable, assuming independence of the parent–child allele transmissions, and incorporating a θ correction to allow for coancestry (recent shared inheritance, due for example to population subdivision) among the parents. These are given again in Table 1.

There may be a tendency for different women who bear children by the same man to have some genetic similarity, for example at genes involved in mate choice. However, the effect of any such tendency is expected to be limited to a few genes and so, when the mothers' profiles are available, we assume that they are independent of whether H_0 or H_1 is true. Under this assumption, when the mothers' genotypes are available, (1) can be rewritten:

$$\mathrm{LR} = \frac{1}{2}\left(1 + \frac{P(\text{offspring genotypes}|\text{maternal genotypes}, H_1, \mathrm{Z} = 1)}{P(\text{offspring genotypes}|\text{maternal genotypes}, H_0)}\right). \quad (2)$$

Although there are many possible combinations for the maternal and offspring genotypes at a locus, there are only a few distinct forms for the LR:

- It can be verified from the available genotypes that the offspring do not share a paternal allele. Then LR = 1/2.
- Both paternal alleles can be identified and they match (Table 2, rows 2–6).
- The paternal allele of only one offspring can be identified and this matches one of the other offspring's alleles (Table 2, rows 7–9).
- Neither paternal allele can be identified, and the offspring have the same, heterozygous genotype (Table 2, row 10).
- Neither paternal allele can be identified, and the offspring are both heterozygous but share only one allele (Table 2, row 11).

Examples of genotypes corresponding to each distinct form of the LR are shown in the first two columns of Table 2, while the corresponding LRs (assuming no coancestry, $\theta = 0$) are shown in the third column.

As an example of the $\theta = 0$ LR calculations, consider the case shown in the third row of Table 2, for which the mothers genotypes are AA and AB, while the two offspring genotypes are both AA. In this case, the denominator of the ratio in (2) is the joint probability of:

- the maternal transmissions of an A allele to both offspring, and
- both offspring receiving an A allele from their unrelated fathers.

The numerator of the ratio also includes the first term above, which therefore cancels. The second term of the denominator is $(P_A)^2$, where P_A is the population frequency of allele A, while for the numerator, the corresponding term is the probability that the ibd paternal allele is A, which is P_A. Thus,

$$\mathrm{LR} = \frac{1}{2}\left(1 + \frac{P_A}{(P_A)^2}\right) = \frac{1}{2}\left(1 + \frac{1}{P_A}\right).$$

The final column of Table 2 shows the θ-adjusted formulas corresponding to the scenario in which all the parents of the two individuals have a common level of coancestry measured by θ. This may apply, for example, when all the parents are drawn from a common subpopulation of the population from which the allele proportions are estimated. We omit details of the derivation, for which see [2,3]. Informally, if the shared paternal allele is A, then the number of A alleles in the mothers' genotypes is indicative of the frequency of allele A in the subpopulation. For example, in the case of Table 2 row 3, discussed above, the mothers' genotypes include three A alleles, suggesting that A is common in the subpopulation so that the observation of paternal A alleles from distinct fathers becomes less surprising than when $\theta = 0$, and the LR is correspondingly reduced.

2.2. Inferences using multi-locus profiles

Assuming independence of the individuals' genotypes at distinct loci, whole-profile LRs are obtained as the product of single-locus values. This independence assumption requires that the loci are unlinked, which may not always hold. We conducted a small simulation study to investigate the effects of weak linkage, and found that for a pair of loci separated by 80 cM, the ratio of the LR ignoring linkage and