

Kinship testing with X-chromosomal markers: Mathematical and statistical issues

Michael Krawczak*

Institut für Medizinische Informatik und Statistik, Christian-Albrechts Universität Kiel, Brunswiker Strasse 10, 24113 Kiel, Germany

Received 23 January 2007; accepted 27 January 2007

Abstract

Use of X-chromosomal markers for kinship testing is meaningful if the identical-by-descent allele sharing probabilities of at least two individuals involved in the case differ under the different hypotheses made about the composite relationships. In this situation, optimal decision making about one or the other hypothesis should be based upon the likelihood ratio of the genotype data obtained. When more than one X-chromosomal marker is being used, this implies that the patterns of linkage and linkage disequilibrium between the respective loci have to be taken into account. Otherwise, the evidence extracted from the data by means of the likelihood ratio may be misleading. Exact likelihood calculations on complex pedigrees can be performed using available software such as, for example, the “LINKAGE” programmes widely used in genetic epidemiology. The required genetic maps can be created using physical location information available in public databases.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Kinship; Likelihood ratio; X chromosome; Haplotype; Linkage; Recombination

1. Introduction

In principle, autosomal genetic markers alone would be sufficient to solve any type of disputed kinship, given that such a case can be solved by means of genetic analysis. In some instances, however, markers located on the sex chromosomes may turn out to be more informative than their autosomal counterparts, and the use of sex-chromosomal markers may therefore substantially improve the efficiency of a kinship test. This gain of evidential power is immediately obvious for Y-chromosomal markers, which allow the confirmation or exclusion of an affiliation to the same male bloodline for even remotely related individuals [1]. Unfortunately, however, Y-chromosomal markers are only capable of providing information about kinship if at least one of the alternative hypotheses about the disputed relationships involves an unbroken chain of male-to-male inheritance.

2. Relative efficiency of X-chromosomal markers

Similarly, albeit in a less pronounced fashion, X-chromosomal markers are particularly informative for inferring

parent–offspring relationships that involve at least one female (i.e. mother–daughter, mother–son, and father–daughter duos). For example, if paternity to a daughter is disputed in a classical trio, the mean exclusion chance of an unrelated male equals

$$\sum_i f_i^3(1 - f_i)^2 + \sum_i f_i(1 - f_i)^3 + \sum_{i < j} f_i f_j (f_i + f_j)(1 - f_i - f_j)^2 \quad (1)$$

for autosomal markers and

$$\sum_i f_i^3(1 - f_i) + \sum_i f_i(1 - f_i)^2 + \sum_{i < j} f_i f_j (f_i + f_j)(1 - f_i - f_j) \quad (2)$$

for X-chromosomal markers [2]. Here, f_i denotes the frequency of the i th allele of the marker in question. The two formulas differ only by the exponent of the last factor in each summation. Since this factor is always smaller in formula (2) than in formula (1), the mean exclusion chance of an X-chromosomal marker is consistently larger than that of an autosomal marker with the same allele frequencies. This discrepancy is illustrated in Fig. 1 for two markers with equally frequent alleles. Obviously, the relative efficiency of an X-chromosomal marker is highest when the level of polymorphism is low (as for SNPs)

* Tel.: +49 431 597 3200; fax: +49 431 597 3193.

E-mail address: krawczak@medinfo.uni-kiel.de.

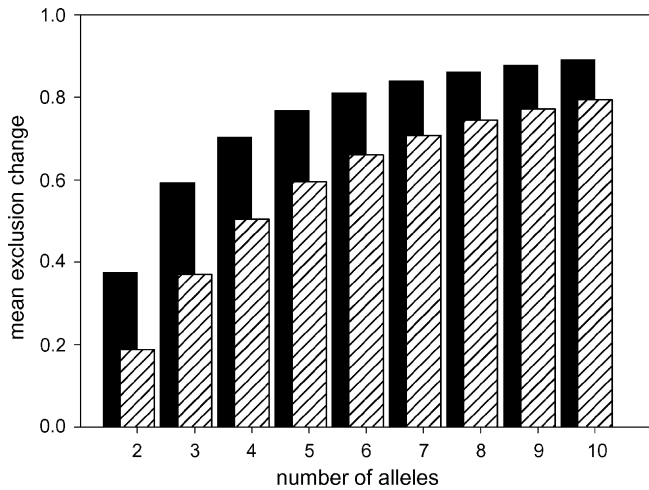


Fig. 1. Mean paternity exclusion chance (MEC) in trios involving a daughter. MEC values are given for autosomal (shaded bars) and X-chromosomal (solid bars) markers with a given number of equally frequent alleles.

and levels off when the degree of polymorphism is high (as for most STRs).

3. A criterion for marker informativity in kinship analyses

The above considerations beg the question: Which formal criterion might we employ to decide whether a certain type of genetic marker (i.e. autosomal, sex-chromosomal, or mitochondrial) could potentially be informative for a given kinship case? In this context, it is worth remembering that each kinship case defines an unambiguous, marker type-specific set H_1, H_2, \dots, H_n of mutually exclusive hypotheses about the so-called ‘identity-by-descent (IBD)’ probabilities of the individuals involved. IBD probability $r_i(k, \phi)$ denotes the probability, under hypothesis H_i , that a pair ϕ of individuals shares k alleles identical by descent ($k = 0, 1, 2$). One plausible criterion for a given marker to be informative for the kinship case in question is that $r_i(k, \phi) \neq r_j(k, \phi)$ for at least one ϕ , at least one k and at least one pair $H_i \neq H_j$. In other words, the marker potentially provides evidence about the case if the IBD probabilities of at least one pair of individuals differ among the different hypotheses made. Although a mathematical proof of this assertion is still pending, it may be surmised that the above condition is both necessary and sufficient. It is almost certainly sufficient because the marker in question would provide information about the relationship between those two individuals who showed differing IBD probabilities under the different hypotheses. It is also most probably necessary because it is hard to imagine a kinship case that can be solved by means of genetic testing, but in which none of the pair-wise degrees of relatedness is disputed.

4. Example 1

Use of the above criterion is exemplified in Fig. 2 for a case involving three individuals, a parent–child duo together with a putative nephew of the parent. If individual 1 is male (Fig. 2a),

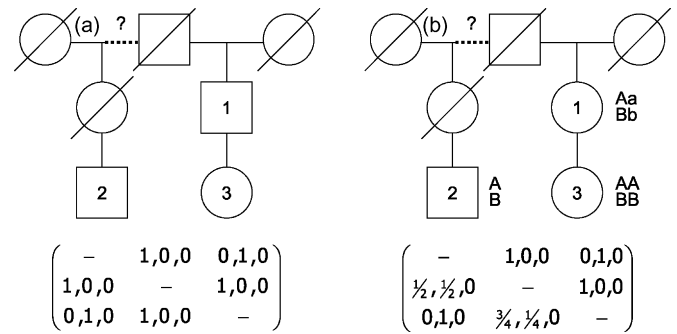


Fig. 2. Kinship cases involving a disputed uncle–nephew (a) and aunt–nephew (b) relationship, respectively. The identity-by-descent probability vectors of an X-chromosomal marker are given below each pedigree, with the bottom left half corresponding to hypothesis H_1 (implied blood relationship) and the top right half corresponding to H_2 (no blood relationship between individuals 1 and 2).

all pair-wise X-chromosomal IBD probability vectors are the same under the two hypotheses H_1 (“nephew–uncle”) and H_2 (“unrelated”). If individual 1 is female, however, the IBD probability vectors of individuals 1 and 2, and of individuals 1 and 3, differ between H_1 and H_2 (Fig. 2b), which implies that an X-chromosomal marker could potentially provide information about the case.

At first glance, it may appear as if genotyping individual 3 for the case depicted in Fig. 2 would be superfluous because this individual is only related to individual 2 via individual 1. This conjecture would indeed be correct if only a single X-chromosomal marker were to be analysed. However, as we shall see, individual 3 may provide essential information about the case if multiple markers are typed. In such instances, the haplotypic arrangements (i.e. the phase) of the used markers, and their possible recombination, have to be taken into account for the necessary likelihood calculation (it should be noted that a similar requirement may also arise when linked autosomal markers are being used for kinship testing).

5. Likelihood calculations on pedigrees

It has become generally accepted in forensics that the best way to include genetic evidence into the decision making process about a case is via the use of the likelihood ratio:

$$LR = \frac{P_i(D)}{P_j(D)} \tag{3}$$

where D denotes data (i.e. genotypes) and $P_i(D)$ is the probability of observing the data given that the i th hypothesis under consideration is correct [3,4]. Here, D usually stands for the observed genotypes of a subset of individuals who are most critical for inference making. However, D can also include other phenotypic data such as enzyme concentrations, blood groups or physical appearance. In this more general form, $P_i(D)$ translates into

$$P_i(D) = \sum_G P(D|G)P_i(G) \tag{4}$$

where summation is over all possible composite genotypes G for the individuals of interest. For most practical purposes, the probability of the data given the genotype, $P(D|G)$, can be assumed to be the product of individual-specific terms $P(D_j|G_j)$, i.e.

$$P(D|G) = \prod_j P(D_j|G_j) \quad (5)$$

where D_j and G_j are the data and genotype belonging to the j th individual, respectively. Similarly, $P_i(G)$ factorises into

$$P_i(G) = \prod_j P_i^*(G_j) \quad (6)$$

where $P_i^*(G_j)$ equals either the population frequency of G_j , if the parents of the j th individual are not involved, or the conditional probability of G_j given the genotypes of the parents, as specified by H_i . Thus, formula (6) reflects the fact that all individuals involved in a kinship case can be distinguished as being either internal family members or so-called ‘founders’, and that the genotypes of all founders are independent of each other. Furthermore, it follows from the Mendelian inheritance of the genetic markers used in forensic practice that the genotypes of internal members are also independent of each other, given that the genotypes of the parents of these individuals are known.

The necessity of correctly taking linkage and linkage disequilibrium between markers into consideration in a kinship case is illustrated by the $\log_{10}(\text{LR})$ curves given in Fig. 3. Assuming the genotypes of Fig. 2b, three different sets of haplotype frequencies f_x were used for two diallelic markers with equally frequent alleles: the solid curve shows $\log_{10}(\text{LR})$ as a function of the recombination fraction θ between the two markers, but in the absence of linkage disequilibrium (i.e. $f_{AB} = f_{Ab} = f_{aB} = f_{ab} = 0.25$); the dotted curve depicts strong linkage disequilibrium between alleles A and B ($f_{AB} = f_{ab} = 0.4$, $f_{Ab} = f_{aB} = 0.1$); the dashed curve corresponds to a strong negative association between A and B ($f_{AB} = f_{ab} = 0.1$,

$f_{Ab} = f_{aB} = 0.4$). Obviously, the three curves differ substantially, particularly for close linkage (i.e. for small values of θ). This discrepancy highlights the importance of using the correct haplotype frequencies and the correct θ in the likelihood calculations. If haplotype AB is rare and linkage is close, the genotype data argue strongly in favour of H_1 . This is because the genotype of individual 3 in this case suggests that individuals 1 and 2 share the rare haplotype AB, which is more likely to have occurred because of common ancestry than by chance alone.

6. The LINKAGE programs as a means of kinship analysis

The above calculations were carried out using the LINKAGE software, a suite of computer programs that were originally developed for use in genetic epidemiology [5]. The component program MLINK performs likelihood calculations on arbitrary pedigrees, although there are some limitations in terms of marker, allele and haplotype number, depending upon the program version. Furthermore, MLINK only allows consideration of one marriage or consanguinity loop in each pedigree. Owing to the nature of the algorithm implemented in MLINK [6], the computation time scales linearly in pedigree size, but exponentially in marker number, which should not normally pose any problems with forensic applications. It has to be cautioned, however, that recent versions of MLINK internally rescale likelihoods to avoid underflow errors, so that some rescaling is required before likelihoods can be directly used in kinship testing. The LINKAGE programs are freely available via the internet at <http://linkage.rockefeller.edu>.

7. Genetic maps

Use of the MLINK program requires that a genetic map of the respective markers is available. This is not an issue for most of the autosomal markers in current forensic use, which are either located on different chromosomes or are so loosely linked that the influence of linkage disequilibrium and a reduced rate of meiotic recombination upon the respective results is negligible. In contrast, forensic markers on the X chromosome are known to cluster so tightly that linkage matters, and that sometimes even linkage disequilibrium has to be taken into account [7]. Unfortunately, comprehensive

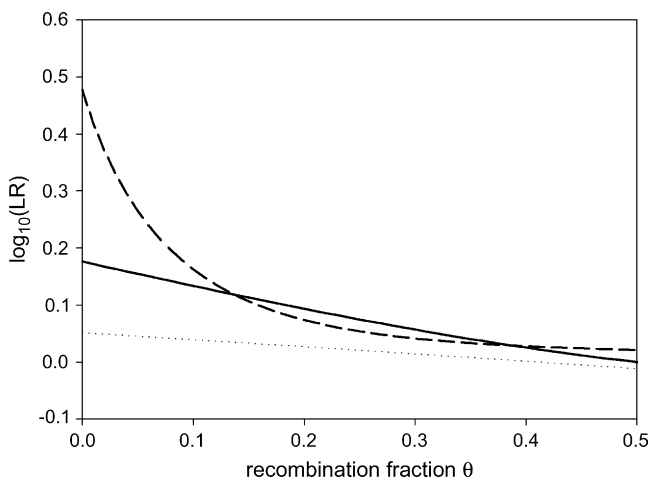


Fig. 3. Evidential power of X-chromosomal markers as a function of linkage and linkage disequilibrium. $\log_{10}(\text{LR})$ values are plotted for the kinship case depicted in Fig. 2. For details: see text.

Table 1
Physical and genetic map of the Mentype[®] Argus X-8 kit

Marker	NCBI (location (Mb))	θ
DXS10135	9.131	–
DXS8378	9.330	0.0020
DXS7132	65.002	0.3358
DXS10074	66.795	0.0176
HPRTB	133.437	0.3681
DXS10101	133.456	0.0002
DXS10134	149.283	0.1357
DXS7423	149.462	0.0018

θ : recombination fraction with marker in previous row, extrapolated from physical marker positions via Haldane’s mapping function (Ref. [8], p. 18).

Table 2
Mentype[®] Argus X-8 genotypes of a kinship case involving a disputed aunt–nephew relationship

Marker	MJ	SJ
DXS10135	21	21–28
DXS8378	10	10–10
DXS7132	13	13–17
DXS10074	17	17–18
HPRTB	12	12–14
DXS10101	29.2	29.2–29.2
DXS10134	36	33–36
DXS7423	14	15–15

genetic maps of the forensic X-chromosomal markers, as determined by comprehensive family analysis, are not available. Alternatively, however, approximate genetic maps can also be created from physical maps using the rule of thumb (Ref. [8], p. 11) that one megabase of DNA sequence corresponds to a genetic distance of one centimorgan (i.e. one expected recombination event per 100 meioses). In this way, the genetic map in Table 1 has been assembled for the markers included in the Mentype[®] Argus X-8 kit. All physical locations were retrieved from NCBI databases (www.ncbi.nlm.nih.gov).

8. Example 2

The Mentype[®] Argus X-8 kit (<http://en.biotype.de>) was recently used in a practical case in which a male client ('MJ') wished to assess whether his mother was the full sister of another woman ('SJ', hypothesis H_1) or was unrelated to that woman (H_2). The respective genotypes (Table 2) revealed a strong similarity between the X chromosomes of MJ and SJ, with one allele being shared identical by state for all but the most distal marker on Xp. Exact quantification of the evidential

power using MLINK and the genetic map in Table 1 yielded a $\log_{10}(\text{LR})$ of 1.628 in favour of H_1 . This result gained further support from the analysis of the 15 autosomal markers included in the Powerplex kit, which yielded a $\log_{10}(\text{LR})$ of 1.921. Taken together the two kits gave a likelihood ratio of 3540 in favour of H_1 , which to all intents solved the case.

Acknowledgement

I am most grateful to Peter JP Croucher, Kiel, for his most helpful comments on the manuscript.

References

- [1] P. Gill, C. Brenner, B. Brinkmann, B. Budowle, A. Carracedo, M.A. Jobling, P. de Knijff, M. Kayser, M. Krawczak, W.R. Mayr, N. Morling, B. Olaisen, V. Pascali, M. Prinz, L. Roewer, P.M. Schneider, A. Sajantila, C. Tyler-Smith, DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, *Forensic Sci. Int.* 124 (2001) 5–10.
- [2] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, *Int. J. Legal Med.* 117 (2003) 67–74.
- [3] B. Devlin, Forensic inference from genetic markers, *Stat. Methods Med. Res.* 2 (1993) 241–262.
- [4] J.S. Buckleton, C.M. Triggs, C. Champod, An extended likelihood ratio framework for interpreting evidence, *Sci. Justice* 46 (2006) 69–78.
- [5] G.M. Lathrop, J.M. Lalouel, C. Julier, J. Ott, Strategies for multilocus linkage analysis in humans, *Proc. Natl. Acad. Sci. U.S.A.* 81 (1984) 3443–3446.
- [6] R.C. Elston, J. Stewart, A general model for the genetic analysis of pedigree data, *Hum. Hered.* 21 (1971) 523–542.
- [7] R. Szibor, S. Hering, E. Kuhlisch, I. Plate, S. Demberger, M. Krawczak, J. Edelmann, Haplotyping of the DXS6801–DXS6809–DXS6789 cluster of STR markers on Xq21 provides a powerful tool for kinship testing, *Int. J. Legal Med.* 119 (2005) 363–369.
- [8] J. Ott, *Analysis of Human Genetic Linkage*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1999.