

The probability of detecting erroneously assigned parentage using co-dominant loci

Mohamed ANALLA ¹, Inmaculada JIMÉNEZ-GAMERO ²,
Ángeles ALONSO-MORAGA ² & Andrés MUÑOZ-SERRANO ²

(Received 02/09/1999 ; Revised 08/11/2000 ; Accepted 08/01/2001)

نسبة تحديد أغلاط الإنتساب الأبوي باستعمال مورثات متعادلة السيطرة

إن تباين الطرق المطروحة حالياً لاحتساب نسبة إلغاء الانتسابات المغلوطة قد يبدو جد محير. هذا العمل يحاول طرح معادلة إجمالية بعد مراعاة جميع الحالات الممكنة. قد يكون النمط الوراثي لأحد الأباء غير معروف. ففي بعض الدراسات، للتأكد من الأب يكون النمط الوراثي للأم غير معروف. فالرفع في هذه الحالات من عدد المورثات المستعملة قد يعوض هذا النقص. لكن معرفة النمط الوراثي للأم يكون ذا أهمية عندما لا يكون الفرد الذي تحت الدرس متشابه الإقتران ويكون نمطه الوراثي مخالف لنمط أمه. من جهة أخرى، يبدو أن عدم وجود توازن هاردي-وينبرغ أو توازن العلاقة بخصوص المورثات المستعملة لتحديد الأبوة يطرح مشكلة قد لا تتوفر على حل سهل و واضح.

الكلمات المفتاحية : النسب المغلوطة - تحديد الأبوة - نسبة نظرية

La probabilité de détection des erreurs d'enregistrement de filiation à travers l'utilisation de loci codominants

La grande variété de formules proposées pour le calcul des probabilités d'exclusion de filiation erronée est vraiment déconcertante. Le présent travail essaie de donner une formule générale en considérant tous les cas possibles. Le génotype du parent non contesté peut ne pas être connu. En effet, dans certaines études d'exclusion du père, aucune information n'est disponible en ce qui concerne le génotype de la mère. Dans ces cas, l'augmentation du nombre de marqueurs utilisés peut compenser ce manque d'information. Cependant, pour que la connaissance du génotype de la mère soit utile, il faut que l'individu dont la filiation est contestée soit hétérozygote et que son génotype soit différent de celui de sa mère. Par ailleurs, l'absence d'équilibre de Hardy-Weinberg ou de liaison, pour les marqueurs utilisés pour la vérification de la filiation, ne semble pas avoir une solution simple et claire.

Mots clés : Généalogie erronée - Contrôle de parenté - Probabilité théorique

The probability of detecting erroneously assigned parentage using co-dominant loci

The variability of formulae nowadays proposed to calculate theoretical probabilities for the exclusion of erroneous parentage could be disconcerting. This work tries to give a general formula considering all possible cases. The genotype of the not contested parent is not always available. Effectively, in some studies of sire exclusion, the information on the genotype of the dam could be lacking. In such cases, increasing the number of the loci used, can balance this lack of information. However, the knowledge of the dam genotype is informative only when the genotype of the individual with contested parentage is heterozygous and different from that of its dam. On the other hand, the lack of Hardy-Weinberg or linkage equilibrium for the loci used for parentage exclusion do not seem to have a simple and clear solution.

Key words : Erroneous genealogy - Parentage control - Theoretical probability

¹ Department of Biology, Faculty of Sciences, Abdelmalek Essaadi University, B.P. 2121, 93002 Tétouan, Morocco,

² Department of Genetics, University of Córdoba, 14005 Córdoba, Spain

^o Corresponding author, e-mail : analla@fst.ac.ma

INTRODUCTION

Several formulae were proposed to calculate the theoretical probability for detecting falsely recorded parents using genetic incompatibilities (Chakraborty & Schull, 1976; Garber & Morris, 1983; Jamieson, 1965 & 1994). Although all these formulae are generally based on allelic frequencies of the putative parents, the discordance between results stemming from different formulae is a bit disappointing. Researchers or technical staff working in parentage exclusion need a unified criterion to apply under rather broad conditions, in order to calculate the required probabilities.

This work aims to discuss some of the already existing formulae and revise all the cases of parentage exclusion, presenting a unified formulation through elementary topics of population genetics.

The theoretical population assumed is one of a large size under both Hardy-Weinberg and linkage equilibria for all the loci considered. The polymorphism used is for autosomal genes, and could be a result of studies via microsattelites (Tautz, 1989), Restriction Length Fragment Polymorphism (RLFP, Botstein *et al.*, 1980), or Single Strand Conformation Polymorphism (SSCP, Orita *et al.*, 1989) or even protein polymorphism via electrophoretic techniques (Smithies, 1955).

Only sire exclusion probability will be explained, since the same formulae will apply to the dam exclusion probability. Albeit, sire exclusion probability will depend on the dam genotype of the individual whose parentage is contested, knowledge of such a genotype will sometimes improve the exclusion process.

Nevertheless, this information is not always available, especially in the case of small ruminants under rather extensive managerial conditions, e.g. sheep and goats in central and south-eastern Spain and northern Morocco.

$$PPX = p_1^2(p_2^2 + p_3^2 + 2p_2p_3) + p_2^2(p_1^2 + p_3^2 + 2p_1p_3) + p_3^2(p_1^2 + p_2^2 + 2p_1p_2) + 2p_1p_2p_3^2 + 2p_2p_3p_1^2 + 2p_1p_3p_2^2$$

Arranging:

$$PX = p_1^2(p_2 + p_3)^2 + p_2^2(p_1 + p_3)^2 + p_3^2(p_1 + p_2)^2 + 2p_1p_2p_3^2 + 2p_1p_3p_2^2 + 2p_2p_3p_1^2$$

DISCUSSION

1. When the dam genotype is not known

The formula of sire exclusion probability, when the dam genotype is not known, was initially proposed by Garber & Morris (1983). Assume a single locus with two alleles: A and B with frequencies p_1 and p_2 , respectively. The genotypic frequencies will be p_1^2 , $2p_1p_2$, and p_2^2 for genotypes AA, AB and BB, respectively, in individuals population, as well as in dams or in sires population. If an individual i is taken at random, its genotype could be:

- AA with probability p_1^2 , therefore, the genotype of sire of i should contain at least one allele A, i. e. possible sires of i should have one of the two genotypes: AA or AB. Then the excluded sires are those with genotype BB. Consequently, the proportion of these excluded sires in the sires' population is $p_1^2p_2^2$.
- BB with probability p_2^2 , therefore, the genotype of sire of i should contain at least one allele B, i. e. possible sires of i should have one of the two genotypes: BB or AB. Then the excluded sires are those with genotype AA. Consequently, the proportion of these excluded sires in the sires' population is $p_2^2p_1^2$.
- AB with probability $2p_1p_2$, therefore, the genotype of sire of i should contain at least one allele A or one allele B, i.e. possible sires of i should have one of the following genotypes: AA, AB or BB. Then no sires would be excluded. Consequently, the proportion of excluded sires is zero.

Then the theoretical probability of sire exclusion (PPX), using one locus with 2 alleles is:

$$PPX = p_1^2p_2^2 + p_2^2p_1^2 + 2p_1p_2 \cdot 0$$

Arranging:

$$PPX = p_1^2(1-p_1)^2 + p_2^2(1-p_2)^2 + 2p_1p_2(1-p_1-p_2)^2$$

When the locus considered is with three alleles: A, B and C, say, with frequencies p_1 , p_2 and p_3 , respectively, the same reasoning, as summarised in table 1, leads to the following expression:

It follows:

$$PPX = p_1^2(1-p_1)^2 + p_2^2(1-p_2)^2 + p_3^2(1-p_3)^2 + 2p_1p_2(1-p_1-p_2)^2 + 2p_1p_3(1-p_1-p_3)^2 + 2p_2p_3(1-p_2-p_3)^2$$

When the locus considered is with four alleles: A, B, C and D, say, with frequencies p_1, p_2, p_3 and p_4 , respectively, the same reasoning, as summarised in table 2, conduces to the following expression:

$$PPX = p_1^2(p_2^2 + p_3^2 + p_4^2 + 2p_2p_3 + 2p_2p_4 + 2p_3p_4) + p_2^2(p_1^2 + p_3^2 + p_4^2 + 2p_1p_3 + 2p_1p_4 + 2p_3p_4) + p_3^2(p_1^2 + p_2^2 + p_4^2 + 2p_1p_2 + 2p_1p_4 + 2p_2p_4) + p_4^2(p_1^2 + p_2^2 + p_3^2 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3) + 2p_1p_2(p_3^2 + p_4^2 + 2p_3p_4) + 2p_1p_3(p_2^2 + p_4^2 + 2p_2p_4) + 2p_1p_4(p_2^2 + p_3^2 + 2p_2p_3) + 2p_2p_3(p_1^2 + p_4^2 + 2p_1p_4) + 2p_2p_4(p_1^2 + p_3^2 + 2p_1p_3) + 2p_3p_4(p_1^2 + p_2^2 + 2p_1p_2)$$

Arranging:

$$PPX = p_1^2(1-p_1)^2 + p_2^2(1-p_2)^2 + p_3^2(1-p_3)^2 + p_4^2(1-p_4)^2 + 2p_1p_2(1-p_1-p_2)^2 + 2p_1p_3(1-p_1-p_3)^2 + 2p_1p_4(1-p_1-p_4)^2 + 2p_2p_3(1-p_2-p_3)^2 + 2p_2p_4(1-p_2-p_4)^2 + 2p_3p_4(1-p_3-p_4)^2$$

It follows that the general formula for sire exclusion using one locus with n alleles, when the dam genotype is not known, is:

$$PPX = \sum_{i=1}^n p_i^2(1-p_i)^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j(1-p_i-p_j)^2$$

2. When the dam genotype is known

In the case of a single locus with two alleles, applying the same conditions as when the dam genotype was not known, if an individual is taken at random from the population, its genotype could be:

- AA with probability p_1^2 . If the dam is AA, this event have a probability p_1 (Table 3), legitimate sires should transmit one A. Then the excluded sires are those with genotype BB. If the dam genotype is AB with probability p_2 , legitimate sires should transmit one A. Then the excluded sires are those with genotype BB. The occurrence of a dam with genotype BB is impossible. Therefore, when the individual genotype is AA, the probability of excluded sires is $p_1^2(p_1+p_2)p_2^2$, which is equal to $p_1^2(1-p_1)^2$. This is the same probability as in the case when the dam genotype was not known.
- BB with probability p_2^2 , therefore, the probability of excluded sires is $p_2^2(p_2+p_1)p_1^2$, which is equal to $p_2^2(1-p_2)^2$; is the same probability as in the case when the dam genotype was not known. Consequently, when the individual is homozygous, knowing the dam genotype do not bring any information to improve sire exclusion probability.

Table 1. Probability of sire exclusion, when ignoring the dam genotype, for an individual I randomly chosen with a probability P(I), using one locus with three alleles: A, B, and C, with frequencies p_1, p_2 and p_3 , respectively

I	P(I)	Possible sires	Excluded sires	Exclusion probability
AA	p_1^2	AA, AB, AC	BB, BC, CC	$p_2^2 + 2p_2p_3 + p_3^2$
BB	p_2^2	AB, BB, BC	AA, AC, CC	$p_1^2 + 2p_1p_3 + p_3^2$
CC	p_3^2	AC, BC, CC	AA, AB, BB	$p_1^2 + 2p_1p_2 + p_2^2$
AB	$2p_1p_2$	AA, AB, AC, BB, BC	CC	p_3^2
AC	$2p_1p_3$	AA, AB, AC, BC, CC	BB	p_2^2
BC	$2p_2p_3$	AB, AC, BB, BC, CC	AA	p_1^2

Table 2. Probability of sire exclusion, when ignoring the dam genotype, for an individual I randomly chosen with a probability P(I), using one locus with four alleles : A, B, C and D with frequencies p_1, p_2, p_3 and p_4 , respectively

I	P(I)	Excluded sires	Exclusion probability
AA	p_1^2	BB, BC, BD, CC, CD, DD	$p_2^2 + 2p_2p_3 + 2p_2p_4 + p_3^2 + 2p_3p_4 + p_4^2$
BB	p_2^2	AA, AC, AD, CC, CD, DD	$p_1^2 + 2p_1p_3 + 2p_1p_4 + p_3^2 + 2p_3p_4 + p_4^2$
CC	p_3^2	AA, AB, AD, BB, BD, DD	$p_1^2 + 2p_1p_2 + 2p_1p_4 + p_2^2 + 2p_2p_4 + p_4^2$
DD	p_4^2	AA, AB, AC, BB, BC, CC	$p_1^2 + 2p_1p_2 + 2p_1p_3 + p_2^2 + 2p_2p_3 + p_3^2$
AB	$2p_1p_2$	CC, CD, DD	$p_3^2 + 2p_3p_4 + p_4^2$
AC	$2p_1p_3$	BB, BD, DD	$p_2^2 + 2p_2p_4 + p_4^2$
AD	$2p_1p_4$	BB, BC, CC	$p_2^2 + 2p_2p_3 + p_3^2$
BC	$2p_2p_3$	AA, AD, DD	$p_1^2 + 2p_1p_4 + p_4^2$
BD	$2p_2p_4$	AA, AC, CC	$p_1^2 + 2p_1p_3 + p_3^2$
CD	$2p_3p_4$	AA, AB, BB	$p_1^2 + 2p_1p_2 + p_2^2$

- AB with probability $2p_1p_2$, therefore, if the dam genotype is also AB, no sires are excluded, as in the case when the dam genotype was not known. However, when the dam genotype is not AB, i.e. dam genotype is different from the individual genotype, further sire could be excluded than in the case when the dam genotype was not known (Table 4).

Table 3. Deduction of dam genotype distribution according the offspring genotype, using one locus with two alleles : A and B with frequencies p_1 and p_2 , respectively

D.m.	A.d.f.g.	A.d.m.g.		D.o.g.*		
		A (p_1)	B (p_2)	AA	AB	BB
AA (p_1^2)	A (p_1^2)	AA ($p_1p_1^2$)	AB ($p_2p_1^2$)	p_1	$0.5 p_1$	0
AB ($2p_1p_2$)	A (p_1p_2)	AA ($p_2p_1^2$)	AB ($p_1p_2^2$)	p_2	0.5	p_1
	B (p_1p_2)	AB ($p_2p_1^2$)	BB ($p_1p_2^2$)			
BB (p_2^2)	B (p_2^2)	AB ($p_1p_2^2$)	BB ($p_2p_2^2$)	0	$0.5 p_2$	p_2

D.g. : Dam genotype ; A.d.f.g. : Allele distribution in female gametes ; A.d.m.g. : Allele distribution in male gametes ; D.o.g. : Distribution of offspring genotype
 * Obtained after arrangement, e.g. if the offspring is AA, the dam is AA with probability $p_1p_1^2/(p_1p_1^2 + p_2p_1^2)$ or AB with probability $p_2p_1^2/(p_1p_1^2 + p_2p_1^2)$.

Table 4. Probability of sire exclusion, when considering M the dam genotype, for an individual I randomly chosen with a probability P(I), using one locus with two alleles : A and B with frequencies p_1 and p_2 , respectively

I	P(I)	M	P(M)	Excluded sires	Exclusion probability
AA	p_1^2	AA	p_1	BB	p_2^2
		AB	p_2	BB	p_2^2
		BB	0	-	-
BB	p_2^2	AA	0	-	-
		AB	p_1	AA	p_1^2
		BB	p_2	AA	p_1^2
AB	$2p_1p_2$	AA	$0.5p_1$	AA	p_1^2
		AB	0.5	None	0
		BB	$0.5p_2$	BB	p_2^2

Thus, summing up, when the individual is homozygous or the individual and his dam have the same genotype, knowing the genotype of the dam does not increase sire exclusion probability. Nevertheless, in the remaining cases, knowledge of the dam genotype gives a valuable information that allows increasing the proportion of falsely assigned sires detected. This could be confirmed

when applying the same reasoning to a locus with 3 or 4 alleles (Tables 5 and 6). Then, the probability of sire exclusion, knowing the genotype of the dam, PPD_X, could be deduced from table 6 as follow:

1. The individual is homozygous for the considered locus, AA, say (first case in Table 6), the term under Exclusion probability (last column in Table 6) reduces to $(1-p_1)^2$, the terms under P(M) (second column in Table 6) sum to 1. Then multiplying by p_1^2 , the probability of AA, we obtain $p_1^2(1-p_1)^2$. Thus, summing for all the other alleles, substituting i for 1, we obtain [Eq.1]:

$$\sum_{i=1}^n p_i^2(1-p_i)^2 \quad \text{[Eq. 1]}$$

This is exactly the same expression as when the dam genotype was not known. Effectively the same sires are excluded in both cases (Tables 2 and 6).

2. The individual is heterozygous for the considered locus, AB, say (fifth case in Table 6), in all the rows of this fifth case the sires with genotype CC, CD or DD are always excluded. These are the same sires which were excluded when the dam genotype was not known (Tables 2 and 6). Their corresponding exclusion probability is $2p_1p_2(p_3^2+2p_3p_4+p_4^2)$, which reduces to $2p_1p_2(1-p_1-p_2)^2$. Nevertheless, when the dam genotype is different from the individual genotype (the dam genotype is not AB), further sires are excluded. Effectively, in second, third and fourth rows of this fifth case, sires with genotype AA, AC or AD are also excluded. Their corresponding exclusion probability is obtained summing all the corresponding terms under P(M) and multiplying this resulting sum times the corresponding terms among those remaining under Exclusion probability, i.e. $2p_1p_2[0.5(p_1+p_3+p_4)(p_1^2+2p_1p_3+2p_1p_4)]$, which reduces to $2p_1p_2[0.5(p_1-p_1p_2)(2-p_1-2p_2)]$. In fifth, sixth and seventh rows of this fifth case, likewise, sires with genotype BB, BC or BD are excluded. Their exclusion probability is, likewise, $2p_1p_2[0.5(p_2-p_2p_1)(2-p_2-2p_1)]$.

Combining All these expressions and summing over all the alleles (substituting i for 1 and j for 2) we obtain equation 2 [Eq. 2].

Then, the general expression of sire exclusion probability using one locus with n alleles, knowing the dam genotype, is [Eq. 3].

Table 5. Probability of sire exclusion, when considering M the dam genotype, for an individual I randomly chosen with a probability P(I), using one locus with three alleles : A, B and C with frequencies p₁, p₂ and p₃, respectively

I	P(I)	M	P(M)	Excluded sires	Exclusion probability
AA	p ₁ ²	AA	p ₁	BB, BC, CC	p ₂ ² +2p ₂ p ₃ +p ₃ ²
		AB	p ₂	BB, BC, CC	p ₂ ² +2p ₂ p ₃ +p ₃ ²
		AC	p ₃	BB, BC, CC	p ₂ ² +2p ₂ p ₃ +p ₃ ²
		BB	0	-	-
		BC	0	-	-
		CC	0	-	-
BB	p ₂ ²	AA	0	-	-
		AB	p ₁	AA, AC, CC	p ₁ ² +2p ₁ p ₃ +p ₃ ²
		AC	0	-	-
		BB	p ₂	AA, AC, CC	p ₁ ² +2p ₁ p ₃ +p ₃ ²
		BC	p ₃	AA, AC, CC	p ₁ ² +2p ₁ p ₃ +p ₃ ²
		CC	0	-	-
CC	p ₃ ²	AA	0	-	-
		AB	0	-	-
		AC	p ₁	AA, AB, BB	p ₁ ² +2p ₁ p ₂ +p ₂ ²
		BB	0	-	-
		BC	p ₂	AA, AB, BB	p ₁ ² +2p ₁ p ₂ +p ₂ ²
		CC	p ₃	AA, AB, BB	p ₁ ² +2p ₁ p ₂ +p ₂ ²
AB	2p ₁ p ₂	AB	0.5(p ₁ +p ₂)	CC	p ₃ ²
		AA	0.5p ₁	AA, AC, CC	p ₁ ² +2p ₁ p ₃ +p ₃ ²
		AC	0.5p ₃	AA, AC, CC	p ₁ ² +2p ₁ p ₃ +p ₃ ²
		BB	0.5p ₂	BB, BC, CC	p ₂ ² +2p ₂ p ₃ +p ₃ ²
		BC	0.5p ₃	BB, BC, CC	p ₂ ² +2p ₂ p ₃ +p ₃ ²
		CC	0	-	-
AC	2p ₁ p ₃	AC	0.5(p ₁ +p ₃)	BB	p ₂ ²
		AA	0.5p ₁	AA, AB, BB	p ₁ ² +2p ₁ p ₂ +p ₂ ²
		AB	0.5p ₂	AA, AB, BB	p ₁ ² +2p ₁ p ₂ +p ₂ ²
		BB	0	-	-
		BC	0.5p ₂	BB, BC, CC	p ₂ ² +2p ₂ p ₃ +p ₃ ²
		CC	0.5p ₃	BB, BC, CC	p ₂ ² +2p ₂ p ₃ +p ₃ ²
BC	2p ₂ p ₃	BC	0.5(p ₂ +p ₃)	AA	p ₁ ²
		BB	0.5p ₂	AA, AB, BB	p ₁ ² +2p ₁ p ₂ +p ₂ ²
		AA	0	-	-
		AB	0.5p ₁	AA, AB, BB	p ₁ ² +2p ₁ p ₂ +p ₂ ²
		AC	0.5p ₁	AA, AC, CC	p ₁ ² +2p ₁ p ₃ +p ₃ ²
		CC	0.5p ₃	AA, AC, CC	p ₁ ² +2p ₁ p ₃ +p ₃ ²

Table 6. Probability of sire exclusion, when considering M the dam genotype, for an individual I randomly chosen with a probability P(I), using one locus with four alleles: A, B, C and D with frequencies p₁, p₂, p₃ and p₄, respectively

I	P(I)	M	P(M)	Excluded sires	Exclusion probability
AA	p ₁ ²	AA	p ₁	BB, BC, BD, CC, CD, DD	p ₂ ² +2p ₂ p ₃ +2p ₂ p ₄ +p ₃ ² +2p ₃ p ₄ +p ₄ ²
		AB	p ₂	BB, BC, BD, CC, CD, DD	p ₂ ² +2p ₂ p ₃ +2p ₂ p ₄ +p ₃ ² +2p ₃ p ₄ +p ₄ ²
		AC	p ₃	BB, BC, BD, CC, CD, DD	p ₂ ² +2p ₂ p ₃ +2p ₂ p ₄ +p ₃ ² +2p ₃ p ₄ +p ₄ ²
		AD	p ₄	BB, BC, BD, CC, CD, DD	p ₂ ² +2p ₂ p ₃ +2p ₂ p ₄ +p ₃ ² +2p ₃ p ₄ +p ₄ ²
BB	p ₂ ²	BB	0	-	-
		BC	0	-	-
		BD	0	-	-
		CC	0	-	-
CC	p ₃ ²	CC	0	-	-
		CD	0	-	-
		DD	0	-	-
		AB	0.5(p ₁ +p ₂)	CC, CD, DD	p ₃ ² +2p ₃ p ₄ +p ₄ ²
DD	p ₄ ²	AA	0.5p ₁	CC, CD, DD, AA, AC, AD	p ₃ ² +2p ₃ p ₄ +p ₄ ² +p ₁ ² +2p ₁ p ₃ +2p ₁ p ₄
		AC	0.5p ₃	CC, CD, DD, AA, AC, AD	p ₃ ² +2p ₃ p ₄ +p ₄ ² +p ₁ ² +2p ₁ p ₃ +2p ₁ p ₄
		AD	0.5p ₄	CC, CD, DD, AA, AC, AD	p ₃ ² +2p ₃ p ₄ +p ₄ ² +p ₁ ² +2p ₁ p ₃ +2p ₁ p ₄
		AB	2p ₁ p ₂	BB	0.5p ₂
AB	2p ₁ p ₂	BC	0.5p ₃	CC, CD, DD, BB, BC, BD	p ₃ ² +2p ₃ p ₄ +p ₄ ² +p ₂ ² +2p ₂ p ₃ +2p ₂ p ₄
		BD	0.5p ₄	CC, CD, DD, BB, BC, BD	p ₃ ² +2p ₃ p ₄ +p ₄ ² +p ₂ ² +2p ₂ p ₃ +2p ₂ p ₄
		CC	0	-	-
		CD	0	-	-
AC	2p ₁ p ₃	DD	0	-	-
		AC	2p ₁ p ₃	.	.
		AD	2p ₁ p ₄	.	.
		BC	2p ₂ p ₃	.	.
BC	2p ₂ p ₃	BD	2p ₂ p ₄	.	.
		CD	2p ₃ p ₄	.	.

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j \left[(1 - p_i - p_j)^2 + 0.5 \left[(p_i - p_i p_j)(2 - p_i - 2p_j) + (p_j - p_j p_i)(2 - p_j - 2p_i) \right] \right] \quad [\text{Eq. 2}]$$

$$\text{PPDX} = \sum_{i=1}^n p_i^2 (1 - p_i)^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j \left[(1 - p_i - p_j)^2 + 0.5 \left[(p_i - p_i p_j)(2 - p_i - 2p_j) + (p_j - p_j p_i)(2 - p_j - 2p_i) \right] \right] \quad [\text{Eq. 3}]$$

It is clear that the knowledge of dam genotype increases the exclusion probability of erroneously assigned sires since the right most term, which does not appear in the expression when the dam genotype was ignored (PPX), is always positive. Nevertheless such an increase is worthwhile only in the case of a locus with a small number of alleles, i.e. a locus with a reduced polymorphism.

Effectively, Table 7 allows a comparison between some values of the probability of sire exclusion, when ignoring or when considering the dam genotype, for an increased number of loci, with 2 to 20 alleles each (the alleles of the same series having the same allelic frequency).

In the case of a locus with two alleles, considering the dam genotype increases the exclusion probability by 50%, whilst, in the case of a locus with 20 alleles, considering the dam genotype increases this probability by about 10% only. Such an increase is further reduced when using more than one locus, e.g. when using five loci with ten alleles each the increase is only 0.4%.

Therefore, as a first consequence of this latter fact, when the dam genotype is not known, increasing the number of loci used should compensate for such a lack of information.

The second consequence is that when the loci are highly polymorph, the knowledge of the individual

and the sire genotype are the only mandatory information

3. More than one locus

Using a single locus could not be efficient enough in order to reach the required exclusion probability. The number of alleles for the used locus could be small, not allowing in this way to reach the required value of exclusion.

Using two or more loci together solves this problem. Such loci should, however, be in linkage equilibrium to ensure that the genotypic frequencies of two loci are therefore independent, and they could be multiplied in order to get the combined genotypic frequencies.

Bearing in mind this condition, and following Boyd (1954), the expression of the global probability of parentage exclusion GPX, using m loci is [Eq. 4]:

$$GPX = 1 - \prod_{i=1}^m (1 - PPX_1) \quad [\text{Eq. 4}]$$

Or when the dam genotype is considered [Eq. 5]:

$$GPX = 1 - \prod_{i=1}^m (1 - PPD\!X_1) \quad [\text{Eq. 5}]$$

Table 7. Some values of the probability of sire exclusion, when the dam genotype is not known (PPX), and when this genotype is known (PPDX), according to the number of loci and alleles used

Alleles*	Loci									
 1 5 10 15 20 10 15 20 10 15
	PPX	PPDX	PPX	PPDX	PPX	PPDX	PPX	PPDX	PPX	PPDX
2	0.1250	0.1875	0.4871	0.6459	0.7369	0.8746	0.8651	0.9556	0.9307	0.9843
4	0.3281	0.5039	0.8631	0.9700	0.9813	0.9902	0.9974	1.0000	0.9996	1.0000
6	0.4861	0.6842	0.9642	0.9954	0.9987	0.9991	1.0000	1.0000	1.0000	1.0000
8	0.5879	0.7434	0.9881	0.9989	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
10	0.6570	0.7947	0.9953	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	0.7066	0.8291	0.9978	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0.7725	0.8723	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0.8146	0.8981	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

* All alleles of the same series with the same frequency

4. When equilibria conditions do not hold

When the population under study is not under Hardy-Weinberg equilibrium, the expressions above can be expressed in terms of genotypic rather than allelic probabilities. The expressions get quite complicated. Furthermore, the inference about the values of exclusion probability obtained will be valid only in the generation in which they were calculated. Chakraborty *et al.* (1988) had tackled the problem when allelic frequencies are different between males and females. This is one case where Hardy-Weinberg equilibrium is not assumed. But the formulae outlined by these authors lay on the fact that both male population and female population are each under Hardy-Weinberg equilibrium. A quite less probable situation (females are in equilibrium, males are in equilibrium but the whole population is not), making, therefore, of little uses the proposed formulae. Usha *et al.* (1995) proposed a quite tedious method (PRASE) to circumvent the lack of both Hardy-Weinberg and linkage equilibria. But, such a formulation could only apply to the generations present in the sample where it was calculated. Because, in such a case no inference to the population could be made as the genotypic frequencies could vary in the next generations. Furthermore, lack of equilibrium could rise in small size samples as a consequence of sampling errors. Nevertheless, a theoretical probability should not be deduced from a small size sample. Because the methodology used (Maximum Likelihood) in order to estimate probabilities from sample observed frequencies gives asymptotically best estimates, i.e. they have small error variance only when the sample size get large.

ACKNOWLEDGEMENTS

First author highly appreciates the financial support of the University of Cordoba during his stay at "Departamento de Genética de la Universidad de Cordoba".

REFERENCES

- Botstein D., White R.L., Skolmic M. & Davis R.W. (1980) Construction of a genetic linkage map using restriction fragment polymorphisms. *American Journal of Human Genetics*. 32: 314-331
- Boyd W.C. 1954. Tables and nomograms for calculating chances of excluding paternity. *American Journal of Human Genetics*. 6: 423-433
- Chakraborty R. & Schull W.J. (1976) A note on the distribution of the number of exclusions to be expected in paternity testing. *American Journal of Human Genetics*. 28: 615-618
- Chakraborty R., Meagher T.R. & Smouse P.E. (1988) Parentage analysis with genetic markers in natural populations. I- The expected proportion of offspring with unambiguous paternity. *Genetics*. 118: 527-536
- Garber R.A. & Morris J.W. (1983) General equation for the average power of exclusion for genetic systems of n codominant alleles in one-parent and in no-parent cases of disputed parentage. *In: Inclusion Probabilities in parentage Testing* (ed. by R.H. Walker), pp. 277-280. American Association of Blood Banks. Arlington, Virginia
- Jamieson A. (1965) The genetics of transferrin in cattle. *Heredity*. 20: 419-441
- Jamieson A. (1994) The effectiveness of using codominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. *Animal Genetics*. 25: 37-44
- Orita H., Suzuki Y., Sekiya T. & Hayachi K. (1989) Rapid and sensitive detection of point mutation using the PCR. *Genomics*. 5: 874-879
- Smithies O. (1955) Zone electrophoresis in starch gels. Group variation in the serum protein of normal human adults. *Biochemical Journal*. 61: 629-641
- Tautz D. (1989) Hypervariability of simple sequences as general source for polymorphic DNA markers. *Nucleic Acids Research*. 17: 6463-6471
- Usha A.P., Simpson S.P. & Williams J.L. (1995) Probability of random sire exclusion using microsatellite marker for parentage verification. *Animal Genetics*. 26: 155-161