Theor Appl Genet (1996) 92:966–975

K. G. Dodds · M. L. Tate · J. C. McEwan
A. M. Crawford

Exclusion probabilities for pedigree testing farm animals

Received: 10 February 1995/Accepted: 6 October 1995

Abstract Pedigree testing, using genetic markers, may be undertaken for a variety of situations, of which the classical paternity testing is only one. This has not always been made clear in the literature. Exclusion probabilities associated with various testing situations, including the use of autosomal or X-linked codominant marker systems with any number of alleles, are presented. These formulae can be used to determine the appropriate exclusion probability for the situation being investigated. One such situation is where sire groups of progeny are to be verified without knowledge of the dams' genotypes, in which case the classical paternity exclusion probability is too high, and if used may result in an optimistic declaration about the progeny that have not been excluded. On the other hand, if mating pairs are known then incorrect progeny can be excluded at a higher rate than suggested by paternity exclusion calculations. The formulae also assist in determining the usefulness of X-linked markers, particularly if the pedigree checks involve progeny of only one sex. A system of notation that is useful for the algebraic manipulation of genetic probabilities, including exclusion probabilities as presented here, is also given.

Key words Parentage tests · Exclusion probabilities · Genetic markers

Introduction

In extensively farmed animals, such as sheep and cattle, accurate pedigree records usually rely on a variety of factors, including secure containment of animals, direct evidence of mating and observation of parturition or suckling activity (Alexander et al. 1983). Genetic gain in animal improvement programmes is sensitive to pedigree misidentifications, particularly where there is extensive use of artificial breeding technologies or the number of offspring are limited (Geldermann et al. 1986). Additionally, for some aspects of animal breeding research a very small number of incorrect relationships could have a large influence on the results, for example when searching for segregation in traits (Elston and Stewart 1971; Kinghorn et al. 1993) or when using linked markers to predict genotypes at specific loci (Goddard 1992).

Hypothesized familial relationships of organisms can be tested using variation in inherited genetic markers, but the power of such tests depends on the exact hypothesis in question. In humans, for example, the relationship of a child to its mother is usually assumed to be correct, and tests are made to determine whether the putative father might be the biological father. If the putative father has a genotype incompatible with the child in question, then he is excluded from paternity. However, if the putative father is compatible there are two possibilities: he may either be the genuine father or an unrelated man who, by chance, has a compatible genotype.

We often wish to know how likely we are to uncover whether a putative pedigree is incorrect; this is termed an exclusion probability. For example, paternity exclusion is the probability of a randomly chosen male being excluded from paternity for a randomly chosen mother-offspring pair. Exclusion probabilities associated with paternity testing have been dealt with extensively (Boyd 1954; Jamieson 1965; Chakraborty et al. 1974; Jamieson 1994).

However, there are situations, particularly in pedigree recording of farmed animals, where we may not be willing to assume that the dam-offspring assignment is unambiguous but that there may be some other relationship we assume (Kashi et al. 1990). Alternatively, we

Communicated by E. J. Eisen

K. G. Dodds (✉) · M. L. Tate · J. C. McEwan
AgResearch Invermay Agricultural Centre, Private Bag 50034, Mosgiel, New Zealand

A.M. Crawford
AgResearch Molecular Biology Unit, Department of Biochemistry, University of Otago, Dunedin, New Zealand

may wish to verify only one of the parents, with the other unknown or not of interest. For example, where only the sire-dam mating assignments are known, we may want to test whether a putative offspring belongs to the parent-pair. Pedigree error rates in farm animals, associated with this and other situations, have been examined using genetic variation (McCoubrey et al. 1983; Tate et al. 1992), although formal derivation and comparison of the probability of exclusion for each situation has not been presented. The fact that different pedigree testing situations exist, with different exclusion probabilities, has not always been clear in the literature. Typically paternity exclusion probabilities have been quoted no matter what the situation (Marklund et al. 1994; Pépin et al. 1995).

In most farm animals, there is now a wide choice of suitable codominantly inherited genetic markers of known chromosomal location (Barendse et al. 1994; Bishop et al. 1994; Rohrer et al. 1994; Burt et al. 1995; Crawford et al. 1995). This technology enables strategic testing to a predetermined probability of exclusion to minimize the confounding effects of pedigree errors.

Some authors have discussed paternity exclusion with dominant systems (Wiener et al. 1930: Boyd 1954; Weiner 1975; Bucher and Elston 1995; Weir 1990) such as the ABO human blood group system. For two alleles, with the dominant allele at frequency p, the paternity exclusion is $p(1-p)^4$, which gives a maximum exclusion of 0.082 when $p = 0.2$ (less than half the best possible codominant two-allele case). As exclusion probabilities are less for dominant than for codominant systems, and because of the availability of codominant markers, we will concern ourselves with codominant systems only.

We derive the exclusion probability formulae for a variety of situations, relevant particularly to farm animals, of which the classical paternity testing is only one. While some of these probabilities have previously been given explicitly (Jamieson 1965; Crawford et al. 1993) or implicitly (Meagher and Thompson 1986), the present paper presents the derivations for the different situations in a unified manner. Both autosomal and homogametic sex-linked codominant marker systems are examined, and the formulae presented allow for any number of alleles in a system. When referring to the sex-linked situation we use mammalian terminology. We also present a system of notation, referred to as the S-notation, that is useful for the algebraic manipulation of genetic probabilities, including exclusion probabilities. We outline the relationships between these probabilities of exclusion and discuss the effectiveness of autosomal or X-linked markers in different test situations.

S-Notation

Definitions

We consider a marker locus with m alleles, with p_i denoting the frequency of A_i the ith allele. We also let

$$S_t = \sum_i p_i^t$$

$$S_{tu} = \sum_{i\neq} \sum_j p_i^t p_j^u$$

$$S_{tuv} = \sum_{i\neq} \sum_{j\neq} \sum_k p_i^t p_j^u p_k^v$$

$$S_{tuvw} = \sum_{i\neq} \sum_{j\neq} \sum_{k\neq} \sum_l p_i^t p_j^u p_k^v p_l^w \quad (1)$$

where i, j, k and l all range from 1 to m; summation subscripts such as $i \neq j \neq k$ indicate that the first sum is over all i, the second is for $j \neq i$ and the third is for $k \neq i, k \neq j$ and so on; and t, u, v and w are arbitrary non-negative integers. Although we deal only with exclusion probabilities, the S-notation and properties (see below) are also useful for other genetic probability calculations.

Because the order of the subscripts in S quantities is irrelevant to the definitions, we will use the convention of ordering the subscripts by size, e.g. we use S_{21} and not S_{12}. Also, because the sums are defined in terms of p subscripts being different, but of any order (i.e. $i < j$ or $i > j$), the probabilities of allele sets need to take this into account, as shown in the following examples. Unless noted otherwise, alleles designated with different subscripts will refer to different alleles.

In this section, we illustrate the notation with examples, along with their interpretation (assuming a population in Hardy-Weinberg equilibrium). The probability of the genotype A_iA_j, $(i \neq j)$ is p_ip_j, and the sum of all such probabilities (i.e. the probability of a heterozygote) is S_{11}. Another example is to calculate the probability that mates are both homozygous, but for different alleles. In this case we want the probability of the pair of genotypes A_iA_i and A_jA_j, $i \neq j$, which is $p_i^2p_j^2$. Summing over all possible combinations of $i \neq j$ gives the probability S_{22}.

Example: Two-allele case

Suppose $m = 2$, with one allele having frequency p and the other $q(= 1-p)$. We find that $S_{11} = 2pq$, which is the probability of an A_1A_2 individual, and $S_{22} = 2p^2q^2$, which corresponds to the first individual being A_1A_1 (with probability p^2) and the mate being $A_2A_2(q^2)$ or vice versa (with the same probability).

Properties

An S quantity with h subscripts is equal to zero if the marker system has fewer than h distinct alleles. This follows from the fact that such an S quantity is defined as a sum over all combinations of sets of h distinct alleles. Some additional properties are as follows:

$$S_0 = m \text{ (the number of distinct alleles)}$$

$$S_1 = 1$$

$$S_{tu} = S_tS_u - S_{(t+u)}$$

$$S_{tuv} = S_{tu}S_v - S_{(t+v)u} - S_{(u+v)t}$$

$$S_{tuvw} = S_{tuv}S_w - S_{(t+w)uv} - S_{(u+w)tv} - S_{(v+w)tu} \quad (2)$$

An example of the derivation for these is as follows:

$$\begin{aligned} S_{tu} &= \sum_{i\neq} \sum_j p_i^t p_j^u \\ &= \sum_{i=1}^{m} p_i^t \left(\sum_{j=1}^{m} p_j^u - p_i^u \right) \\ &= \left(\sum_i p_i^t \right) \left(\sum_j p_j^u \right) - \left(\sum_i p_i^{t+u} \right) \\ &= S_tS_u - S_{(t+u)} \end{aligned}$$

Therefore, equations in terms of general S quantities can be expressed in terms of S quantities with single subscripts (although this may not be the best form for computational accuracy because it involves higher indices in the summations). For example, the probability of a heterozygote is given above as S_{11}, but this can also be expressed as $1-S_2$ (i.e. the probability of not being homozygous). Our other example, S_{22}, is $S_2^2-S_4$, i.e. the probability of two homozygous individuals, less the probability that they are homozygous for the same alleles. Jamieson (1994) also finds that he can express probabilities of interest in terms of the single-subscripted S quantities (in our notation), but derives each specific case.

Example: Two-allele case

The two-allele case has further simplifications. Any S quantity with more than two subscripts is zero. We found above that $S_{11}=2pq$, but from the general results, this also equals $1-S_2$. Now, because subscripts are in non-increasing order,

$$\begin{aligned} S_{tu} &= p^t q^u + p^u q^t \\ &= p^u q^u (p^{t-u} + q^{t-u}) \\ &= \frac{(1-S_2)^u}{2^u} S_{(t-u)} \end{aligned} \tag{3}$$

which expresses a double-subscripted S quantity in terms of S quantities with single subscripts (2 and $t-u$), which are numerically no larger than either of the original subscripts.

We can also express all single-subscripted S quantities in terms of S_2 as follows. Letting $u=1$ in the general result gives $S_{t1}=S_tS_1-S_{(t+1)}=S_t-S_{(t+1)}$; equating this to Eq. 3 and rearranging yields $2S_{(t+1)}=2S_t-(1-S_2)S_{(t-1)}$. As an example we can use these results to simplify S_{22} to $(1-S_2)^2S_0/4=(1-S_2)^2/2=2p^2q^2$ as found previously. These relationships allow comparison of the results in this article with published results for the two-allele case.

Situations

We assume that the parents are randomly sampled from a single population (so that the marker has the same allele frequencies for sires and dams) that is in Hardy-Weinberg equilibrium (or its X-linked analogue, Bulmer 1980) at the locus of interest. In addition, we assume that mating in this population is random, that alleles are transmitted to progeny in a Mendelian fashion, that the progeny are randomly chosen (the locus is unaffected by selection), that there is no mutation at the locus of interest and that genotypes are determined without error.

We examine the situations for autosomal markers listed in Table 1, and for X-linked markers listed in Table 2. For the purposes of calculating exclusion probabilities, pseudoautosomal markers are considered as autosomal, rather than X-linked. Definitions and derivations of the exclusion probabilities listed in these tables follows.

Autosonal markers

Paternity or maternity exclusion

For autosomal markers, the sex of the parent in question is irrelevant, so we consider only paternity exclusion; the maternity exclusion probability will be the same. The general formula for this case of a genetic system controlled by a number of autosomal codominant alleles is given by Jamieson (1965), and derivations can be found in Selvin (1980) or Weir (1990; Table 6.10). The formula given by these authors and the corresponding formula in our notation are

$$\begin{aligned} Q_1 &= \sum_i p_i(1-p_i)^2 - \frac{1}{2}\sum_{i\neq}\sum_j p_i^2 p_j^2 (4-3p_i-3p_j) \\ &= S_1 - 2S_2 + S_3 - 2S_{22} + \tfrac{3}{2}S_{32} + \tfrac{3}{2}S_{32} \\ &= 1 - 2S_2 + S_3 - 2S_{22} + 3S_{32} \\ &= 1 - 2S_2 + S_3 + 2S_4 - 2S_2^2 - 3S_5 + 3S_3S_2 \end{aligned} \tag{4}$$

The calculations are repeated here for completeness, with the relevant exclusions shown in Table 3. The entries for the exclusion probability term column are

Table 1 Situation investigated for autosomal codominant markers and definitions of Q_1, Q_2 and Q_3[a]

Situation	Sire	Dam	Offspring	Exclusion probability[b]
Sire, dam known	T	Y	T	Q_1
Dam, sire known	Y	T	Y	Q_1
Sire, dam unknown	Y	N	T	Q_2
Dam, sire unknown	N	Y	T	Q_2
Offspring	Y	Y	T	Q_3

[a] Y = known individual, relationship with other Ys known, N = individual unknown or not genotyped, T = individual whose relationship is being tested. Where there is only one Y individual, the roles of Y and T can be interchanged
[b] See text

Table 2 Situation investigated for X-linked codominant markers and definitions of Q_4, Q_5 and Q_6[a]

[a] Y = known individual, relationship with other Ys known, N = individual unknown or not genotyped, T = individual whose relationship is being tested. Where there is only one Y individual, the roles of Y and T can be interchanged
[b] See text

Situation	Sire	Dam	Son	Daughter	Exclusion probability[b]
Sire-son, dam known	T	Y	Y		0
Sire-daughter, dam known	T	Y		Y	Q_4
Dam-son, sire known	Y	T	Y		Q_5
Dam-daughter, sire known	Y	T		Y	Q_5
Sire-son, dam unknown	Y	N	T		0
Sire-daughter, dam unknown	Y	N		T	Q_5
Dam-son, sire unknown	N	Y	T		Q_5
Dam-daughter, sire unknown	N	Y		T	Q_2
Parent-son	Y	Y	T		Q_5
Parent-daughter	Y	Y		T	Q_6

Table 3 Paternity exclusion for an autosomal marker and a known mother-offspring pair (derivation of Q_1)

Dam		Offspring		Sire		Exclusion probability term
Genotype	Probability	Genotype	Probability	Included genotypes	Excluded probability	
A_iA_i	p_i^2	A_iA_i	p_i	A_iA_i, A_iA_j	$(1-p_i)^2$	$S_3-2S_4+S_5$
A_iA_i	p_i^2	A_iA_j	p_j	A_iA_j, A_jA_j, A_jA_k	$(1-p_j)^2$	$S_{21}-2S_{22}+S_{32}$
A_iA_j	p_ip_j	A_iA_i	p_i	A_iA_i, A_iA_j, A_iA_k	$(1-p_i)^2$	$S_{21}-2S_{31}+S_{41}$
A_iA_j	p_ip_j	A_iA_j	$(p_i+p_j)/2$	$A_iA_i, A_iA_j, A_iA_k,$ A_jA_j, A_jA_k	$(1-p_i-p_j)^2$	$S_{21}-2S_{31}-2S_{22}$ $+2S_{32}+S_{41}+S_{32}$
A_iA_j	p_ip_j	A_iA_k	p_k	$A_iA_k, A_jA_k, A_kA_k, A_kA_l$	$(1-p_k)^2$	$S_{111}-2S_{211}+S_{311}$

calculated by multiplying the other probabilities in that row and summing over the distinct subscripts. For example, the first entry is

$$\sum_i p_i^2 \times p_i \times (1-p_i)^2 = \sum_i p_i^3 - 2p_i^4 + p_i^5 \qquad (5)$$
$$= S_3 - 2S_4 + S_5$$

Adding the exclusion probability terms over the possible dam-offspring combinations, and using Eq. 2 gives

$$\begin{aligned} Q_1 &= (S_3 - 2S_4 + S_5) + (S_2 - S_3 - 2S_2^2 + 2S_4 + S_3S_2 \\ &\quad - S_5) + (S_2 - 3S_3 + 3S_4 - S_5) \\ &\quad + (S_2 - 3S_3 + 5S_4 - 4S_5 - 2S_2^2 + 3S_3S_2) \\ &\quad + (1 - 5S_2 + 7S_3 - 6S_4 + 2S_5 + 2S_2^2 - S_3S_2) \\ &= 1 - 2S_2 + S_3 + 2S_4 - 2S_2^2 - 3S_5 + 3S_3S_2 \end{aligned} \qquad (7)$$

as before. This latter form has also been given by Jamieson (1994) (although not in the S-notation). The interpretation of the excluded probability for the sire given in the first row of Table 3 is that neither of his alleles are A_i. The same result can be obtained by specifying the excluded probability as 1 minus the probability of the genotypes listed in the 'Included genotypes' column, e.g. by setting the second last column of the first row in Table 3 to $1-p_i^2-2p_ip_j$. In this case we obtain

$$\begin{aligned} S_3 - S_5 - 2S_{41} &= S_3 - S_5 - 2S_4 + 2S_5 \\ &= S_3 - 2S_4 + S_5 \end{aligned}$$

which is the same as the first bracketed term in Eq. 7. Similar equivalence hold for the other rows in this table, and for other such tables that we will investigate.

Paternity or maternity exclusion, other parent unknown

We refer to this as the single parent-offspring exclusion. This has been given by Crawford et al. (1993) and is repeated here for completeness and to show its representation in the S-notation. The exclusions are listed in Table 4. Adding over possible sire genotypes gives

$$Q_2 = 1 - 4S_2 + 4S_3 - 3S_4 + 2S_2^2 \qquad (8)$$

Table 4 Paternity exclusion for an autosomal marker and unknown dam (derivation of Q_2)

Sire		Offspring		Exclusion probability term
Genotype	Probability	Included genotypes	Excluded probability	
A_iA_i	p_i^2	A_iA_i, A_iA_j	$(1-p_i)^2$	$S_2-2S_3+S_4$
A_iA_j	p_ip_j	$A_iA_i, A_iA_j, A_iA_k,$ A_jA_j, A_jA_k	$(1-p_i-p_j)^2$	$S_{11}-2S_{21}-2S_{21}$ $+S_{31}+S_{31}+2S_{22}$

Table 5 Mating types and excluded offspring for a parent pair (derivation of Q_3)

Parents		Offspring		Exclusion probability term
Genotype	Probability	Included genotypes	Excluded probability	
$A_iA_i \times A_iA_i$	p_i^4	A_iA_i	$1-p_i^2$	S_4-S_6
$A_iA_i \times A_iA_j$	$4p_i^3p_j$	A_iA_i, A_iA_j	$1-p_i^2-2p_ip_j$	$4(S_{31}-S_{51}-2S_{42})$
$A_iA_i \times A_jA_j$	$p_i^2p_j^2$	A_iA_j	$1-2p_ip_j$	$S_{22}-2S_{33}$
$A_iA_i \times A_jA_k$	$2p_i^2p_jp_k$	A_iA_j, A_iA_k	$1-2p_i(p_j+p_k)$	$2(S_{211}-2S_{321}-2S_{321})$
$A_iA_j \times A_iA_j$	$2p_i^2p_j^2$	A_iA_i, A_iA_j, A_jA_j	$1-(p_i+p_j)^2$	$2(S_{22}-S_{42}-2S_{33}-S_{42})$
$A_iA_j \times A_iA_k$	$4p_i^2p_jp_k$	$A_iA_i, A_iA_j, A_iA_k, A_jA_k$	$1-p_i^2-2p_ip_j-2p_ip_k-2p_jp_k$	$4(S_{211}-S_{411}-2S_{321}-2S_{321}-2S_{222})$
$A_iA_j \times A_kA_l$	$p_ip_jp_kp_l$	$A_iA_k, A_iA_l, A_jA_k, A_jA_l$	$1-2(p_i+p_j)(p_k+p_l)$	$S_{1111}-2S_{2211}-2S_{2211}-2S_{2211}-2S_{2211}$

Offspring exclusion

This refers to the case where the matings are known, and therefore the possible parent pairs, but the offspring have not been observed at birth, and so the assignment to parent pairs is in doubt. To derive the exclusion probability we need to consider the seven possible mating types (Jamieson 1965). These are shown in Table 5, along with the excluded types.

Adding up the exclusion probabilities, weighted by the probability of the mating type gives:

$$Q_3 = 1 + 4S_4 - 4S_5 - 3S_6 - 8S_2^2 + 2S_3^2 + 8S_3S_2 \qquad (9)$$

X-linked markers

Paternity exclusion for sons

Because a male does not pass his X chromosome to a son, he cannot be excluded as the sire using an X-linked marker, i.e. the exclusion probability is zero.

Paternity exclusion for daughters

For each dam-offspring possibility, the sire alleles that can be excluded are the same as for the autosomal case. The table becomes that shown in Table 6, and we see that the values in the second last column of this table are square roots of the corresponding values in Table 3, because there is now only one allele to consider for each sire. The exclusion probability is

$$Q_4 = 1 - S_2 + S_4 - S_2^2 \qquad (10)$$

Maternity exclusion for sons

The (known) genotype of the sire is irrelevant here, as he does not pass his X chromosome to a son. The relevant exclusions are in Table 7. The exclusion probability is

$$Q_5 = 1 - 2S_2 + S_3 \qquad (11)$$

Maternity exclusion for daughters

The relevant exclusions are in Table 8. These are similar to the autosomal maternity (or paternity) exclusion where the known parent is homozygous. The exclusion probability is the same as that for the previous situation, i.e.

$$1 - 2S_2 + S_3 = Q_5$$

Table 6 Paternity exclusion for a X-linked marker and a known dam-daughter pair (derivation of Q_4)

Dam		Daughter		Sire		Exclusion probability term
Genotype	Probability	Genotype	Probability	Included genotypes	Excluded probability	
A_iA_i	p_i^2	A_iA_i	p_i	A_i	$1-p_i$	S_3-S_4
A_iA_i	p_i^2	A_iA_j	p_j	A_j	$1-p_j$	$S_{21}-S_{22}$
A_iA_j	p_ip_j	A_iA_i	p_i	A_i	$1-p_i$	$S_{21}-S_{31}$
A_iA_j	p_ip_j	A_iA_j	$(p_i+p_j)/2$	A_i, A_j	$1-p_i-p_j$	$S_{21}-S_{31}-S_{22}$
A_iA_j	p_ip_j	A_iA_k	p_k	A_k	$1-p_k$	$S_{111}-S_{211}$

Table 7 Maternity exclusion for a X-linked marker and a known sire-son pair (derivation of Q_5)

Sire		Son		Dam		Exclusion probability term
Genotype	Probability	Genotype	Probability	Included genotypes	Excluded probability	
Any	1	A_i	p_i	A_iA_i, A_iA_j	$1-p_i^2-2p_ip_j$	$S_1-S_3-2S_{21}$

Table 8 Maternity exclusion for a X-linked marker and a known sire-daughter pair (alternative derivation of Q_5)

Sire		Daughter		Dam		Exclusion probability term
Genotype	Probability	Genotype	Probability	Included genotypes	Excluded probability	
A_i	p_i	A_iA_i	p_i	A_iA_i, A_iA_j	$(1-p_i)^2$	$S_2-2S_3+S_4$
A_i	p_i	A_iA_j	p_j	A_iA_j, A_jA_j, A_jA_k	$(1-p_j)^2$	$S_{11}-2S_{21}+S_{31})$

Sire-son exclusion (unknown dam)

As in the paternity exclusion for sons case we cannot exclude a sire-son relationship, so the exclusion probability is zero.

Sire-daughter exclusion (unknown dam)

The relevant exclusion is in Table 9. The exclusion probability is the same as that for maternity exclusion for a X-linked marker and a known sire-offspring, i.e.

$$1 - 2S_2 + S_3 = Q_5$$

Dam-son exclusion (unknown sire)

This is the same exclusion as the maternity exclusion for sons, because the genotype of the sire was irrelevant there, and so lack of knowledge of this genotype will not affect the probability.

Dam-daughter exclusion (unknown sire)

This is the same situation as the autosomal case (although in that case it applied to any sex of parent and any sex of offspring) because both dam and daughter have two alleles at the X-linked marker.

Parent-son exclusion

The genotype of the sire is irrelevant, so the exclusion is as for dam-son X-linked exclusions.

Parent-daughter exclusion

The relevant exclusions are in Table 10. The exclusions are similar to the autosomal case with a homozygous sire. The exclusion probability is

$$Q_6 = 1 + 2S_4 - S_5 - 4S_2^2 + 2S_3S_2 \qquad (12)$$

Discussion

Comparison of exclusion probabilities

Comparisons of various exclusion probabilities at different allele frequencies (Figs. 1 and 2) allow several generalizations to be made. As noted by Selvin (1980) and Weir (1990), paternity exclusion for a marker with m alleles is greatest when they are at equal frequency. Similar arguments show that this is true for all exclusion probabilities presented; Fig. 1 demonstrates this for the two-allele case. The appendix shows that $Q_6 > (Q_4, Q_3) > Q_5 > Q_1 > Q_2$ for any polymorphic system. This allows us to draw some general conclusions. For an autosomal marker (1) it is always easier to detect incorrect offspring assignments (i.e. when mating pairs are known) than other types of exclusion and (2) paternity (or maternity) exclusion is greater with the other parent known than unknown, as expected.

The comparisons also show that for an autosomal and an X-linked marker with the same allele frequencies, the X-linked marker can provide a greater probability of exclusion, depending on the sex of the putative offspring. For daughters the X-linked marker always has a greater exclusion probability, no matter what situation is being tested (except for maternity testing without knowledge

Table 9 Paternity exclusion for a X-linked marker for daughters with dam unknown (alternative derivation of Q_5)

Sire		Daughter		Exclusion probability term
Genotype	Probability	Included genotype	Excluded probability	
A_i	p_i	A_iA_i, A_iA_j	$(1-p_i)^2$	$S_1 - 2S_2 + S_3$

Table 10 Mating types for an X-linked marker and excluded offspring for a parent pair (derivation of Q_6)

Parents		Daughter		Exclusion probability term
Genotype	Probability	Included genotypes	Excluded probability	
$A_i \times A_iA_i$	p_i^3	A_iA_i	$1-p_i^2$	$S_3 - S_5$
$A_i \times A_iA_j$	$2p_i^2p_j$	A_iA_i, A_iA_j	$1-p_i^2-2p_ip_j$	$2(S_{21} - S_{41} - 2S_{32})$
$A_i \times A_jA_j$	$p_ip_j^2$	A_iA_j	$1-2p_ip_j$	$S_{21} - 2S_{32}$
$A_i \times A_jA_k$	$p_ip_jp_k$	A_iA_j, A_iA_k	$1-2p_i(p_j+p_k)$	$S_{111} - 4S_{221}$

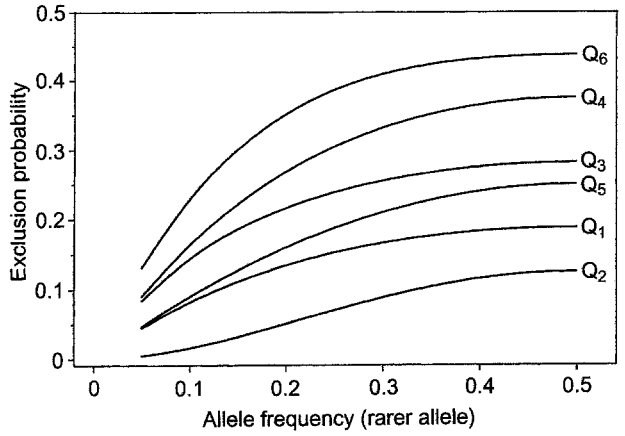

Fig. 1 Exclusion probabilities for markers with two alleles. See text for explanation of Q_1–Q_6

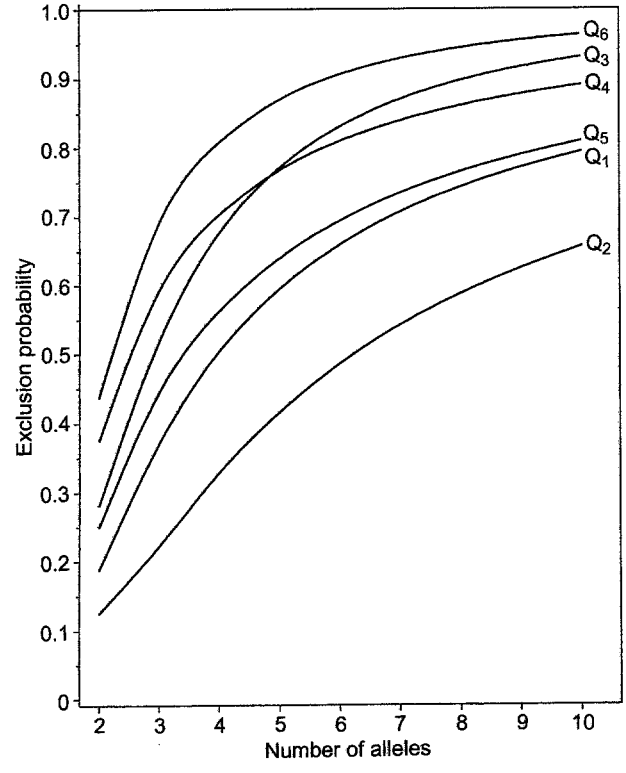

Fig. 2 Exclusion probabilities for markers with alleles at equal frequency. See text for explanation of Q_1–Q_6

of the sire where the autosomal and X-linked markers have the same exclusion probability). For sons, the autosomal marker has a higher exclusion probability except for maternity testing, in which case the X-linked marker would be better.

These comparisons demonstrate that for situations in which only female offspring are of interest, such as progeny testing diary bulls (Van Vleck 1970), addition of an X-linked rather than an autosomal marker to a marker panel for pedigree testing would be much more efficient. Figure 2 shows that in this situation, with unknown dams, an X-linked marker with five equally frequent alleles gives about the same level of exclusion as an autosomal marker with ten equally frequent alleles (compare Q_5 with Q_2). The obvious drawback of using X-linked markers, however, is that paternity testing cannot be carried out for sons (zero exclusion probability).

The results of exclusion probability calculations for autosomal situations where one parent is unknown or where the mates are assumed known are also found in Thompson (1986), but no formulae are given. The first line (labelled 'Parental exclusion probability') in her Table 13 corresponds to the use of our Q_2; however the results in her Table 17 ('Parent-pair exclusion probability') do not match the use of our Q_3. Meagher and Thompson (1986) provide a formula for these exclusion probabilities in very general terms. In their notation paternity (maternity) exclusion is given by $1 - P(NQQ|QU)$, single parent-offspring exclusion is given by $1 - P(NQU|QQ)$ and offspring exclusion is given by $1 - P(NQQ|QQ)$. The use of the formulae given here matches their calculations for two codominant autosomal loci in their Table 1. Thompson (1986) also discusses situations such as hypothesized full-sib relationships with no information on the parents. Such relationships cannot be excluded on the basis of genetic data, and so there is not a corresponding exclusion probability.

It is interesting to note that Q_4 is the same as the polymorphic information content (PIC) of Botstein et al. (1990) because in the notation used here the PIC is $1 - S_2 - S_{22} = 1 - S_2 + S_4 - S_2^2$. The PIC is defined to be the probability of knowing which of a parent's alleles is passed to an offspring.

An application

We now examine the use of one of these exclusion probabilities in farm pedigree testing situations. Crawford et al. (1993) examined the recorded pedigrees of 195 lambs and both of their putative parents with the marker MAF36. The marker detected one incorrect pedigree. If we assume that the recorded matings are correct, then we wish to know the probability of detecting an incorrect assignment to a parent-pair. The use of the lambs' allele frequencies for this marker (0.003, 0.074, 0.112, 0.036, 0.112, 0.464, 0.153, 0.038, 0.008, 0.073) with Q_3 suggests that 73.5% of incorrectly assigned lambs will be detected. Another way of determining how likely an exclusion is is to remove the only excluded assignment (mating pair plus putative lamb) from the dataset, and then compare each mating pair's genotypes with those of all the lambs of other parents. The average proportion of these which can be excluded is an estimate of the chance of detection. These calculations suggest that 74.3% of errors should be detected, close to the theoretical value given above. This result is shown in

Table 11 Comparison of probabilities of detecting incorrectly assigned offspring of known mating pairs for flocks and markers described by Crawford et al. (1993). The appropriate exclusion probabilities (Q_3), calculated using lamb's allele frequencies or weighted parents' allele frequencies, are compared with the average proportion of the lambs of other parents which can be excluded

Flock	Marker	Q_3		Comparison of mates and non-progeny
		Lambs[a]	Parents[b]	
Stud	MAF36	0.756	0.742	0.740
Stud	MAF209	0.602	0.590	0.618
Parasite Flock A	Transferrin	0.722	0.716	0.726
Parasite Flock A	Plasminogen Antigen	0.336	0.324	0.343
Parasite Flock A	Vitamin D binding protein	0.222	0.224	0.278
Parasite Flock B	MAF36	0.735	0.763	0.743

[a] Q_3 calculated using observed allele frequency in lambs
[b] Q_3 calculated using weighted observed allele frequency in parents, as described in the text

Table 11 (where the flock is described as Parasite Flock B) along with a comparison using several of the other markers and flocks from Crawford et al. (1993). The use of these two methods can be seen to generally produce similar results.

In these examples, the allele frequencies could have been calculated in other ways, such as by taking the mean of the allele frequencies of the sires and dams, weighted by the number of offspring they have. This might help account for a relatively small number of sires for the dataset. When this is done for MAF36 in Parasite Flock B, the theoretical exclusion rate becomes 76.3%. Other examples using data from Crawford et al. (1993) are also shown in Table 11.

Multiple markers

Most applications of pedigree testing will use more than one marker. If the set of markers is independently segregating, then the probability of being compatible at all loci is the product of the probabilities of being compatible at each locus (Boyd 1954; Weir 1990). For example, if Q_{1k} is the paternity exclusion for the kth locus, then the combined paternity exclusion probability is

$$Q_1 = 1 - \prod_k (1 - Q_{1k})$$

Jamieson (1994) presents a table to illustrate that for the same total number of alleles, exclusion based on multiple polymorphic loci is higher than when based on a single locus.

It may be desirable to require exclusion at more than one locus to reduce the effect of possible genotyping errors or mutation (Chakraborty et al. 1974; Chakraborty and Schull 1976) or of unknown null alleles (Pemberton et al. 1995). In this case, the multiple locus exclusion probability will be reduced by the probability of exactly one exclusion, so that the paternity exclusion probability becomes

$$Q_1 = 1 - \prod_k (1 - Q_{1k}) - \sum_k Q_{1k} \prod_{l \neq k} (1 - Q_{1l})$$

(Chakraborty et al. 1974). For example, if we have loci with two, three and four equally frequent alleles, then the paternity exclusion probabilities requiring one and more than one exclusion are 0.746 and 0.281 respectively. With more markers it is likely that this difference will decrease (see Chakraborty et al. 1974 for another example).

Violation of assumptions

The formulae presented depend on a number of assumptions, many of which may not be strictly true for some farmed animal populations. The previous section has shown how to reduce the chance of false exclusions due to mutation and genotyping errors. We have also given an application in sheep pedigrees, where Q_3 appears to be consistent with a method which is applicable without the assumptions. These examples have also shown that the assumption of the sires and dams being random members of the same population does not appear to be too critical in practice. This is supported by McCoubrey et al. (1983) who found little difference between using sire allele frequencies weighted or unweighted by the number of their offspring when fitting likelihood models to estimate error rates. One should use care, however, in choosing allele frequency estimates for these calculations, as they may vary between different breeds or subpopulations.

There may be instances where the assumptions are known to be violated to such as extent that the formulae will be in serious error, and their application in these situations would be unwise. If the sire and dam populations are quite distinct, formulae that allow different allele frequencies for the sire and dam populations could be developed, as has been done for paternity exclusion

Table A1 Representation of exclusion probability differences as sums of positive S quantities. The first column displays the inequality being investigated. The second column displays the difference between the exclusion probabilities, expressed in terms of single-subscripted S quantities. The third column shows the same difference, expressed as the sum of S quantities. S quantities with five subscripts are defined analogously to those with fewer subscripts [as in Eq. 1] and may be represented as a linear combination of S quantities with four or fewer subscripts following the pattern shown in Eq. 2

Inequality	Difference	
	Form 1	Form 2
$Q_6 > Q_4$	$S_4 - S_5 + S_2 - 3S_2^2 + 2S_3S_2$	$S_{2111} + 3S_{311} + S_{41}$
$Q_6 > Q_3$	$3S_6 + 3S_5 - 2S_4 + 4S_2^2 - 2S_3^2 - 6S_3S_2$	$S_{411} + 2S_{321} + 4S_{221} + 3S_{42} + S_{41}$
$Q_4 > Q_5$	$S_4 - S_3 + S_2 - S_2^2$	$S_{211} + S_{31}$
$Q_3 > Q_5$	$4S_4 - 4S_5 - 3S_6 - S_3 + 2S_2 - 8S_2^2 + 8S_3S_2 + 2S_3^2$	$S_{32111} + 4S_{22111} + 2S_{21111} + 2S_{4211} + 3S_{3311} + S_{3221} + 12S_{3211} + 7S_{3111} + 6S_{2221} + S_{521} + S_{511} + 5S_{431} + 3S_{421} + 5S_{411} + 4S_{331} + S_{61} + S_{43}$
$Q_5 > Q_1$	$3S_5 - 2S_4 + 2S_2^2 - 3S_3S_2$	$2S_{221} + S_{32}$
$Q_1 > Q_2$	$-3S_5 + 5S_4 - 3S_3 + 2S_2 - 4S_2^2 + 3S_3S_2$	$2S_{2111} + 3S_{311} + 2S_{221} + S_{41}$

by Chakraborty et al. (1988). The population may be known to be highly inbred, or conversely, the parents or the progeny may have been produced by an outcross. The number of possible sires may be limited, rather than a random sample from the populations. For these cases, formulae could also be derived in a similar manner to those of this paper, incorporating inbreeding as well as allele frequency effects on the genotype frequencies or by taking the exactly family structure into account. Alternatively, the method used by Usha et al. (1995) for paternity exclusions, using genotype rather than allele frequencies, could be adapted to the various pedigree testing situations. MacCluer and Schull's (1963) calculations of paternity exclusions, for case where the biological or putative parents are related, could be extended to other pedigree testing situations. Finally, the probabilities could be in serious error if the same markers are being used simultaneously for pedigree checking and marker-assisted selection, and further studies are needed to quantify such effects.

With the exception of extreme situations, where gross violations of the assumptions have occurred, the exclusion probabilities presented here will be of practical use for the variety of situations examined. They will aid investigators who are designing pedigree testing systems for farmed animals, although it would be prudent to use conservative exclusion levels to offset violations of the assumptions that may occur in practice.

Acknowledgements We wish to thank our AgResearch colleagues, and two anonymous reviewers, for helpful suggestions.

Appendix

Exclusion probability inequalities

To show that the one exclusion probability is at least as large as another, it is sufficient to show that their difference (largest minus smallest) can be written as the sum of S quantities. As each of these is the sum of non-negative numbers (products and powers of allele frequencies), the sum of such quantities is also non-negative. Table A1 shows that this can be done for the inequalities referred to in the text of this article. In addition, because each row of the last column in the table contains a double-subscripted S quantity, the differences are strictly positive if there are at least two alleles in the system.

References

Alexander G, Stevens D, Mottershead B (1983) Problems in the accurate recording of lambing data. Aust J Exp Agric Anim Husb 23:361–368

Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM, Clayton D, Li L, Neibergs HL, Zhang N, Grosse WM, Weiss J, Creighton P, McCarthy F, Ron M, Teale AJ, Fries R, McGraw RA, Moore SS, Georges M, Soller M, Womack JE, Hetzel DJS (1994) A genetic linkage map of the boving genome. Nat Genet 6:227–235

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins GA, Toldo SS, Fries R, Grosz MD, Yoo, JY, Beattie CW (1994) A genetic linkage map for cattle. Genetics 136:619–639

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32:314–331

Boyd WC (1954) Tables and nomograms for calculating chances of excluding paternity. Am J Hum Genet 6:426–433

Bucher KD, Elston RC (1975) Estimation of nonpaternity for X-linked trait. Am J Hum Genet 27:689–690

Bulmer MG (1980) The mathematical theory of quantitative genetics. Clarendon Press, Oxford

Burt DW, Bumstead N, Bitgood JJ, Ponce De Leon FA, Crittenden LB (1995) Chicken genome mapping: a new era in avian genetics. Trends Genet 11:190–194

Chakraborty R, Schull WJ (1976) A note on the distribution of the number of exclusions to be expected in paternity testing. Am J Hum Genet 28:615–618

Chakraborty R, Shaw M, Schull WJ (1974) Exclusion of paternity: the current state of the art. Am J Hum Genet 26:477–488

Chakraborty R, Meagher TR, Smouse PE (1988) Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118:527–536

Crawford AM, Tate ML, McEwan JC, Kumaramanickavel G, McEwan KM, Dodds KG, Swarbrick PA, Thompson P (1993) How reliable are sheep pedigrees? Proc N Z Soc Anim Prod 53:363–366

Crawford AM, Dodds KG, Ede AJ, Pierson CA, Montgomery GW, Garmonsway HG, Beattie AE, Davies K, Maddox JF, Kappes

SW, Stone RT, Nguyen TC, Penty JM, Lord EA, Broom JE, Buitkamp J, Schwaiger W, Epplen JT, Matthew P, Matthews ME, Hulme DJ, Beh KJ, McGraw RA, Beattie CW (1995) An autosomal genetic linkage map of the sheep genome. Genetics 140:703–724

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Geldermann H, Pieper U, Weber WE (1986) Effect of misidentification on the estimation of breeding value and heritability in cattle. J Anim Sci 63:1759–1768

Goodard ME (1992) A mixed model for analyses of data on multiple genetic markers. Theor Appl Genet 83:878–886

Jamieson A (1956) The genetics of transferrins in cattle. Heredity 20:419–441

Jamieson A (1994) The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. Anim Genet 25S1:37–44

Kashi Y, Lipkin E, Darvasi A, Nave A, Gruenbaum Y, Beckmann JS, Soller M (1990) Parentage identification in bovine using "Deoxyribonucleic Acid Fingerprints". J Dairy Sci 73:3306–3311

Kinghorn BP, Kennedy BW, Smith C (1993) A method of screening for genes of major effect. Genetics 134:351–360

MacCluer JW, Schull WJ (1963) On the estimation of the frequency of nonpaternity. Am J Hum Genet 15:191–202

Marklund S, Ellegren H, Eriksson S, Sandberg K, Andersson L (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. Anim Genet 25:19–23

McCoubrey CM, Sales DI, Archibald AL (1983) Testing genetic models in populations which contain pedigree errors. Anim Blood Groups Biochem Genet 14:257–268

Meagher TR, Thompson EA (1986) The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor Pop Biol 29:87–106

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1955) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. Mol Ecol 4:249–252

Pépin L, Amigues Y, Lépingle A, Berthier JL, Bensaid A, Vaiman D (1995) Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity 74:53–61

Rohrer GA, Alexander LJ, Keele JW, Smith TP, Beattie CW (1994) A microsatellite linkage map of the porcine genome. Genetics 136:231–245

Selvin S (1980) Probability of nonpaternity determined by multiple allele codominant systems. Am J Hum Genet 32:276–278

Tate ML, Dodds KG, Thomas KJ, McEwan KM (1992) Genetic polymorphism of plasminogen and vitamin D binding protein in red deer, (*Cervus elaphus* L.). Anim Genet 23:303–313

Thompson EA (1986) Pedigree analysis in human genetics. John Hopkins University Press, Baltimore

Usha AP, Simpson SP, Williams JL (1995) Probability of random sire exclusion using microsatellite markers for parentage verification. Anim Genet 26:155–161

Van Vleck LD (1970) Misidentification and sire evaluation. J Dairy Sci 53:1697–1702

Weir BS (1990) Genetic data analysis. Sinauer, Sunderland, Mass.

Wiener AS (1975) Chances of proving nonpaternity with tests for a sex-linked trait. Am J Hum Genet 27:243–245

Wiener AS, Lederer M, Polayes SH (1930) Studies in isohemagglutination. IV. On the chances of proving non-paternity; with special reference to the blood groups. J Immunol 19:259–282