Sib-parentage testing using molecular markers when parents are unknown

D. García*, C. Carleos†, D. Parra* and J. Cañón*

*Laboratorio de Genética, Facultad de Veterinaria, Universidad Complutense de Madrid, Madrid, Spain. †Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Oviedo, Oviedo, Spain

Summary

The formulae for computing the so-called Sib Index using codominant alleles for (1) full-sib and (2) half-sib parentage are given. Hypothesis testing is based on the distribution of conditional likelihood ratio or Bayes’ factor. Thresholds for rejecting the null hypothesis and \( P \)-values were obtained in function of the number of alleles and their frequency distributions. Simulations showed that a relatively low number of marker systems (e.g. 20) are enough to accept the hypothesis of sib parentage with a reasonable power for usual significance levels, but that a higher number would be necessary if full-sib against half-sib parentage is the contrast to be carried out. The effect of sampling variation on the allele frequencies on power calculations is also analysed.

Keywords Bayes’ factor, molecular markers, sib parentage test.

Introduction

There is a need to check the correct paternity relationships to be used for predicting the genetic merit of the individuals included in the breeding programme through the numerator relationship matrix (Henderson 1976), especially for farmed animal species. It is well known that errors in paternity assignment delay genetic progress, whose magnitude under certain circumstances reaches that of the amount of paternity errors in the pedigree (Ron et al. 1996). While in these farmed species paternity testing is generally concerned with the exclusion of paternity, in others the requirement is to establish family relationships for legal, social or medical reasons (Pena & Chakraborty 1994). Exclusions of paternity are irrefutable and the power of a set of genetic markers to exclude is systematically computed into the exclusion probability, which depends on allelic frequencies in the population (Jamieson 1994; Jamieson & Taylor 1997). The availability for most domestic and many wild animal species of a large number of highly informative DNA markers and their utility for checking parentage have increased the possibilities (Goodnight & Queller 1999; Ritland 2000; Fiumera & Asmussen 2001). In particular, showing whether two individuals are sibs when no parental information is accessible is one of the most frequent questions asked for at animal genetic services or forensic laboratories. It is clear that under the situation, in which exclusion of paternity is the goal, the acceptance of the exclusion is irrefutable because, assuming no mutations and no laboratory errors, a descendent must carry parental alleles. However, proof of sibships depends on statistical inference. This paper presents formulae to compute Sib and Half-sib Indexes using codominant markers when parent information is unknown and their distributions under different amounts of marker information.

The conditional likelihood ratio

The conditional likelihood ratio (LR) or Bayes factor has traditionally been considered as the way to evaluate the evidence in paternity disputes (Aitken 1997). It is the ratio of two probabilities, the probability of \( G \) (the genotype of the individuals) when \( S \) (both individuals are sib related) is true \( [\Pr(G|S)] \) and the probability of \( G \) when \( S \) is false \( [\Pr(G|\neg S)] \). The likelihood ratio takes values between 0 and \( \infty \), while its logarithm, which receives the name of the weight of the evidence (Good 1950), takes values on \( [-\infty, \infty] \). It has been

Address for correspondence

Javier Cañón, Laboratorio de Genética,
Facultad de Veterinaria, Universidad Complutense de Madrid,
28040 Madrid, Spain.
E-mail: jcanon@vet.ucm.es

Accepted for publication 29 March 2002

© 2002 International Society for Animal Genetics, Animal Genetics, 33, 364–371
argued in favour of the use of the logarithm of the likelihood ratio that the results it produces are in accordance with the weighing of evidence in the scale of justice.

The hypotheses that we wish to test are: (a) S, both individuals are sibs, (b) I, the individuals are unrelated random individuals from the referenced population, (c) HS, both individuals are half-sibs. Individuals are genotyped and the result is G. Large LR values argue in favour of sib-ships, whereas small values argue against it.

In the context of this paper the probability ratio is called Sib Index or Half-sib Index.

**Calculation of the Sib Index for full-sib testing**

Formally, our interest is to calculate the joint probability of two genotypes conditioned to the fact that the individuals carrying them are sibs, and compare it with the probability of those genotypes when the individuals are not sibs. In order to do that, and assuming that we are dealing with several loci, we will show first how to do it with a particular locus.

For a given locus $A$, let $A_1A_j$ and $A_kA_l$ be the genotypes for $A$ of the two individuals, and denote by $s$ the event of them being indeed whole sibs.

We can define the Sib Index for locus $A$ as

$$SI_A = \frac{P[A_1A_j \wedge A_kA_l | s]}{P[A_1A_j \wedge A_kA_l | \neg s]}.$$  

The probability in the denominator equals, under Hardy–Weinberg equilibrium $(2 - \delta_i)(2 - \delta_j)p_j p_k p_l$, where $p_m$ stands for the frequency of allele $m$ in the population, $m = 1, \ldots, n$, $n$ being the number of alleles in locus $A$, and $\delta_{mk}$ is the well-known Kronecker Delta, which equals 1 if $m = n$ and 0 otherwise. This result is obvious because both individuals are independent in this situation.

To obtain the probability in the numerator, we need a little algebra. We know that $P[A_1A_j \wedge A_kA_l | s] = P[A_1A_j | s] P[A_kA_l | s]$. Now, $s$ gives no information for the calculation of $P[A_kA_l]$, hence $P[A_1A_j | s] = (2 - \delta_j)p_j p_l$ — remember we are assuming HW equilibrium.

We are therefore interested in obtaining the probability of the genotype of one sib conditioned to the fact that it is indeed a sib of the other and to the genotype of the latter. Now,

$$P[A_1A_j | s \wedge A_kA_l] = \sum_{x, y, z, d = 1}^n P[A_1A_j \wedge GP = (A_xA_y \times A_zA_l) | s \wedge A_kA_l],$$

where GP is the genotype of a hypothetical parental mating. But

$$\sum_{x, y, z, d = 1}^n P[A_1A_j \wedge GP = (A_xA_y \times A_zA_l) | s \wedge A_kA_l] = \sum_{x, y, z, d = 1}^n P[A_1A_j | GP = (A_xA_y \times A_zA_l) \wedge s \wedge A_kA_l].$$

The probability in the denominator equals, under Hardy–Weinberg equilibrium $(2 - \delta_i)(2 - \delta_j)p_j p_k p_l$, where $p_m$ stands for the frequency of allele $m$ in the population, $m = 1, \ldots, n$, $n$ being the number of alleles in locus $A$, and $\delta_{mk}$ is the well-known Kronecker Delta, which equals 1 if $m = n$ and 0 otherwise. This result is obvious because both individuals are independent in this situation.

To obtain the probability in the numerator, we need a little algebra. We know that $P[A_1A_j \wedge A_kA_l | s] = P[A_1A_j | s] P[A_kA_l | s]$. Now, $s$ gives no information for the calculation of $P[A_kA_l]$, hence $P[A_1A_j | s] = (2 - \delta_j)p_j p_l$ — remember we are assuming HW equilibrium.

We are therefore interested in obtaining the probability of the genotype of one sib conditioned to the fact that it is indeed a sib of the other and to the genotype of the latter. Now,

$$P[A_1A_j | s \wedge A_kA_l] = \sum_{x, y, z, d = 1}^n P[A_1A_j \wedge GP = (A_xA_y \times A_zA_l) | s \wedge A_kA_l],$$

where GP is the genotype of a hypothetical parental mating. But

$$\sum_{x, y, z, d = 1}^n P[A_1A_j \wedge GP = (A_xA_y \times A_zA_l) | s \wedge A_kA_l] = \sum_{x, y, z, d = 1}^n P[A_1A_j | GP = (A_xA_y \times A_zA_l) \wedge s \wedge A_kA_l].$$

The sum is taken through all the possible matings for a locus with $n$ alleles, but, as $P[GP = (A_xA_y \times A_zA_l) | s \wedge A_kA_l] \neq 0$ only for those matings in which one parent carries at least one $A_i$ allele and the other carries at least one $A_j$ allele, it reduces to

$$\sum_{x, y, z, d = 1}^n P[A_1A_j | GP = (A_xA_y \times A_zA_l) \wedge s \wedge A_kA_l]$$

$$P[GP = (A_xA_y \times A_zA_l) | s \wedge A_kA_l].$$

Now, once the alleles $A_i$ and $A_j$ are fixed in the genotypes of the parents, the only stochastic variation is due to $A_x$ and $A_y$, so

$$P[GP = (A_xA_y | A_zA_l) | s \wedge A_kA_l] = p_x p_y,$$

and therefore

$$P[A_1A_j | s \wedge A_kA_l]$$

$$= \sum_{x, y, z, d = 1}^n p_x p_y P[A_xA_y | GP = (A_xA_y \times A_zA_l) \wedge s \wedge A_kA_l].$$

For example, consider a locus $A$ with three alleles $A_1$, $A_2$, and $A_3$, and suppose that the reference individual is $A_1A_2$, his putative sib $A_1A_3$ and the alleles have frequencies of $p_1$, $p_2$, and $p_3$, Then the possible matings to originate an $A_1A_2$ individual are

$$A_1A_1 \times A_1A_3$$

and thus

$$P[A_1A_1 | s \wedge A_1A_2] = 0.5 p_1 p_3 + 0.25 p_1 p_2 + 0.25 p_3^2$$

$$+ 0.25 p_1 p_3$$

$$= 0.5 p_1 p_3 + 0.25 p_1 (p_2 + p_3)$$

$$= 0.5 p_1 p_3 + 0.25 p_1.$$

It can be easily shown that, in general,

$$P[A_1A_1 | s \wedge A_1A_i] = 0.25 (2 - \delta_i)(2 - \delta_j)p_i p_j + \delta_k (1 - \delta_i)p_k + \delta_i \delta_j$$

$\forall i, j, k, l \in \{1, \ldots, n\}$, $n$ being the number of alleles in the locus.

This formula was obtained by trying to provide a unique expression for the explicit, separate formulae deduced for
Table 1 The nine possible cases for \( P(A_i A_j \mid s \land A_i A_j) \). In rows the possibilities for the individual in the condition; in columns the possibilities for the individual whose genotype’s probability is calculated.

<table>
<thead>
<tr>
<th>Homozygous sharing two alleles ((A_i A_j))</th>
<th>Homozygous sharing one allele ((A_i A_j))</th>
<th>Homozygous sharing zero alleles ((A_i A_j))</th>
<th>Heterozygous sharing two alleles ((A_i A_j))</th>
<th>Heterozygous sharing one allele ((A_i A_j))</th>
<th>Heterozygous sharing zero alleles ((A_i A_j))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homozygous ((A_i A_j)) (0.25(1 + p)^2)</td>
<td>0.25(p_i^2)</td>
<td>0.5((p_i p_j + p_i))</td>
<td>0.5(p_i p_j + 0.5(1 + p_i + p_j))</td>
<td>0.5(p_i p_j)</td>
<td></td>
</tr>
<tr>
<td>Heterozygous ((A_i A_j)) (0.25(p_i + p_j)^2)</td>
<td>0.25(p_i^2)</td>
<td>0.5((p_i p_j + p_i))</td>
<td>0.5(p_i p_j + 0.5(1 + p_i + p_j))</td>
<td>0.5(p_i p_j)</td>
<td></td>
</tr>
</tbody>
</table>

Note that although this rationale is based on the calculation of a conditional probability on the genotype of one of the putative sibs in order to obtain the joint probability of the two putative sibs’ genotype, the choice of the genotype upon which the probability is to be conditioned has no influence on the final result.

**Calculation of the Sib Index for half-sib testing**

Although with some variations that we will discuss later on, the situation is fairly analogous to the full-sib case. Assuming the same locus notation as in the previous section, we look for the value of the quotient \( p[A_i A_j \mid hs \land A_i A_j] \), which will be called the Half-sib Index for locus \( A \), and denoted as \( HSI_1 \), and where \( hs \) denotes the event of the two individuals being half-sibs. For the same reasons as above, we can restrict ourselves to calculate \( p[A_i A_j \mid hs \land A_i A_j] \). Now, instead of expanding this probability by intersecting with the genotypes of the parents, we will do it only with the genotype of the father, denoted as \( GF \), because both individuals share the same father, but not necessarily the same mother. Therefore,

\[
P[A_i A_j \mid hs \land A_i A_j] = \sum_{x=1}^{n} P[A_k A_l \land GF = A_x A_y \mid hs \land A_i A_j].
\]

Now,

\[
\sum_{x=1}^{n} P[A_k A_l \land GF = A_x A_y \mid hs \land A_i A_j] = \sum_{x=1}^{n} P[A_k A_l \mid GF = A_x A_y \land hs \land A_i A_j] P[GF = A_x A_y \mid hs \land A_i A_j] = \sum_{x=1}^{n} P[A_k A_l \mid GF = A_x A_y \land hs \land A_i A_j] P[GF = A_x A_y \mid hs \land A_i A_j] + (1 - \delta_i) P[A_k A_l \mid GF = A_x A_y \land hs \land A_i A_j] P[GF = A_x A_y \mid hs \land A_i A_j] = \sum_{x=1}^{n} 0.5(1 + \delta_i) P[A_k A_l \mid GF = A_x A_y \land hs \land A_i A_j] + (1 - \delta_i) P[A_k A_l \mid GF = A_x A_y \land hs \land A_i A_j].
\]

Consider the simplest example, in which one individual bears genotype \( A_1 A_2 \), and the other \( A_1 A_4 \), all the alleles

Figure 1 Empirical probability density functions of the log-Sib Index (up) and log-Half-sib Index (down) under the two hypotheses tested in each (null hypothesis on left). In the middle, empirical pdf of the Log-PR of being full siblings \((H_o)\) vs. being half-sibs. All of them obtained for a random set of frequencies for 20 markers with five alleles each and \(10^6\) simulations.
being different from each other, we try to find the probability of genotype \( A_1 A_2 \) conditioned to the fact that a half-sib of him carries \( A_1 A_2 \). Then

\[
P[A_1 A_4 | hs \wedge A_1 A_2] = 0.5p_3 [P[A_1 A_4 | GF = A_1 A_1] + P[A_1 A_4 | GF = A_2 A_2]] + 0.5p_4 [P[A_1 A_4 | GF = A_1 A_4] + P[A_1 A_4 | GF = A_2 A_4]].
\]

Now, \( P[A_1 A_4 | GF = A_1 A_1] \) is the probability of the individual carrying allele \( A_1 \) and the father passing \( A_1 \), which equals \( 1/2p_4 \). Therefore, \( P[A_1 A_4 | hs \wedge A_1 A_2] = 0.5p_3[1/2p_4 + 1/2p_4] + 1/2p_4[1/2p_3 + 1/2p_3] = p_3p_4 \).

As in the previous situation, we can build up a table with all the possible cases, such as Table 2.

As before, it is easy to show that the probabilities in each row sum one, as expected from a well defined probability measure.

Again, a more general formula can be induced from the table:

\[
P(A_i A_j | hs \wedge A_i A_j) = 0.25(2(2 - \delta_2)p_i p_j + \delta_2(1 + \delta_2)p_i + \delta_2(1 - \delta_2)p_j),
\]

and the same comments of the previous case regarding alleles ordering in the haplotypes apply here.

Thus, the Half-sib Index for locus \( A \) is defined as

\[
HSI_A = \frac{2(2 - \delta_2)p_i p_j + \delta_2(1 + \delta_2)p_i + \delta_2(1 - \delta_2)p_j}{4(2 - \delta_2)p_i p_j},
\]

and the joint Half-sib Index for multiple and independent loci is

\[
HSI = \sum_{m=1}^{M} HSI_m.
\]

**Hypothesis testing**

Simulation routines were developed to study the performance of the test under several situations. Empirical power values and probability density functions were obtained for different marker configurations, allele frequency situations and significance levels. Allele frequencies were simulated for populations with 10, 20 and 30 markers, all of them having the same number of alleles, either 5 or 10. The sets of frequencies were generated in three different ways: random, uniformly with slight variations and uniformly with slight variations except but some extreme frequency alleles. Uniform with slight variations means here that if the number of alleles for a certain locus is \( n \), each allele is present not with frequency \( 1/n \), but \( 1/n \pm 0.01 \) \( (f_1 = 1/n + 0.01, f_2 = 1/n - 0.01, \text{and so on}, \text{with } f_n = 1/n \text{ if } n \text{ is even and } f_n = 1/n \text{ if } n \text{ is odd}) \). Uniform with slight variations but with some extreme frequency alleles means the same as above except that 20% of the loci involved are considered to have extreme frequency alleles, that is, all the alleles but one are present with frequency 0.01. The power was obtained for three significance levels: 0.001, 0.0001 and 0.00001. To obtain the power of one test for a set of allele frequencies, 100 000 pairs of individuals were simulated with those frequencies under null and another 100 000 under alternate hypotheses. Formulae in Table 1 or 2, depending on the contrast, were applied to the 100 000 pairs on each hypothesis, and thus 200 000 values of the statistic were obtained. 100 000 under each hypothesis. For a significance level \( \alpha \), the \( (1 - \alpha) \times 100 000 \)th highest value of the 100 000 \( H_0 \) values was taken as the rejection threshold, and so the power was obtained as the proportion of \( H_1 \) values above that threshold. Probability density values were empirically approximated by taking histogram values as pdf values, with histogram values calculated for 1000 narrow intervals. The method was applied as well in a real Irish-Setter dog breed population, to test the parentage of two putative full sibs. Allelic frequencies were estimated from a total of 64 individuals.

Tables 3–5 show the figures for the different scenarios. As expected, the increase of information from markers results in an increase in power. Reasonable values are obtained for full-sib parentage determination for a non-rare situation of 20 markers with five alleles each. If the number of markers grows up to 30, maintaining the plausible value of five alleles per marker locus, it makes the test show good levels of power, even for a significance level as low as \( 10^{-4} \). Raising the number of alleles per locus would keep the power high even for \( \alpha = 10^{-5} \). As depicted in Fig. 2, in the perhaps overly optimistic situation of having 10 alleles per locus for a total of 30 markers, the density functions of the natural logarithm of

<table>
<thead>
<tr>
<th></th>
<th>Homozygous sharing two alleles ((A_i A_i))</th>
<th>Homozygous sharing one allele ((A_i A_j))</th>
<th>Homozygous sharing zero alleles ((A_j A_j))</th>
<th>Heterozygous sharing two alleles ((A_i A_j))</th>
<th>Heterozygous sharing one allele ((A_i A_j))</th>
<th>Heterozygous sharing zero alleles ((A_j A_j))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homozygous ((A_i A_i))</td>
<td>(0.5p_i(1 + p_j))</td>
<td>(0.5p_j^2)</td>
<td>(0.5p_i(1 + 2p_j))</td>
<td>(0.5(2p_i p_j + p_j))</td>
<td>(p_j p_i)</td>
<td>(0.5(2p_i p_j + 0.5(p_i + p_j)))</td>
</tr>
<tr>
<td>Heterozygous ((A_i A_j))</td>
<td>(0.25p_i(1 + 2p_j))</td>
<td>(0.5p_i^2)</td>
<td>(0.52p_i p_j + 0.5(p_i + p_j))</td>
<td>(0.5(2p_i p_j + 0.5p_i))</td>
<td>(p_j p_i)</td>
<td>(0.5(2p_i p_j + 0.5p_i))</td>
</tr>
</tbody>
</table>

© 2002 International Society for Animal Genetics, Animal Genetics, 33, 364–371
Table 3 Power values for independent vs. full-sibs hypothesis testing.

<table>
<thead>
<tr>
<th></th>
<th>10 markers</th>
<th></th>
<th>20 markers</th>
<th></th>
<th>30 markers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 al/mark</td>
<td></td>
<td></td>
<td>5 al/mark</td>
<td></td>
<td>5 al/mark</td>
</tr>
<tr>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
</tr>
<tr>
<td>Random</td>
<td>0.44536</td>
<td>0.22934</td>
<td>0.10157</td>
<td>0.79854</td>
<td>0.61631</td>
</tr>
<tr>
<td>Ps-uniform</td>
<td>0.53228</td>
<td>0.30110</td>
<td>0.14205</td>
<td>0.85603</td>
<td>0.69940</td>
</tr>
<tr>
<td>Ps-unf w/ext</td>
<td>0.43149</td>
<td>0.21368</td>
<td>0.09171</td>
<td>0.77649</td>
<td>0.58916</td>
</tr>
</tbody>
</table>

Table 4 Power values for independent vs. half-sibs hypothesis testing.

<table>
<thead>
<tr>
<th></th>
<th>10 markers</th>
<th></th>
<th>20 markers</th>
<th></th>
<th>30 markers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 al/mark</td>
<td></td>
<td></td>
<td>5 al/mark</td>
<td></td>
<td>5 al/mark</td>
</tr>
<tr>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
</tr>
<tr>
<td>Random</td>
<td>0.05431</td>
<td>0.01266</td>
<td>0.00274</td>
<td>0.16089</td>
<td>0.05476</td>
</tr>
<tr>
<td>Ps-uniform</td>
<td>0.05494</td>
<td>0.01323</td>
<td>0.00315</td>
<td>0.18322</td>
<td>0.05651</td>
</tr>
<tr>
<td>Ps-unf w/ext</td>
<td>0.05051</td>
<td>0.01444</td>
<td>0.00328</td>
<td>0.15367</td>
<td>0.05228</td>
</tr>
</tbody>
</table>

Table 5 Power values for full-sibs vs. half-sibs hypothesis testing.

<table>
<thead>
<tr>
<th></th>
<th>10 markers</th>
<th></th>
<th>20 markers</th>
<th></th>
<th>30 markers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 al/mark</td>
<td></td>
<td></td>
<td>5 al/mark</td>
<td></td>
<td>5 al/mark</td>
</tr>
<tr>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
<td>x = 10^{-3}</td>
<td>x = 10^{-4}</td>
</tr>
<tr>
<td>Random</td>
<td>0.05166</td>
<td>0.00818</td>
<td>0.01595</td>
<td>0.08489</td>
<td>0.01600</td>
</tr>
<tr>
<td>Ps-uniform</td>
<td>0.06028</td>
<td>0.01067</td>
<td>0.00165</td>
<td>0.09356</td>
<td>0.02013</td>
</tr>
<tr>
<td>Ps-unf w/ext</td>
<td>0.04208</td>
<td>0.00727</td>
<td>0.00146</td>
<td>0.06632</td>
<td>0.01209</td>
</tr>
</tbody>
</table>

the test under the null (independent individuals) and alternate (full-sibs) hypotheses have almost no intersection, which makes the contrast have very little error. Results also suggest that the test performs better for uniform allelic frequency distributions than for the cases in which alleles with extreme frequencies are involved.

On the other hand, power values for half-sib testage, both vs. independency and vs. full-sibship are not very encouraging, because for optimum conditions of having 30 markers with 10 alleles each and the lowest significance level, power gets to about 75% in the best cases. This outcome was expected, as testing for half-sibship with no parent information implies allowing for different genotypes on the non-shared parent, so there is a loss of information which lowers the power of the test. Figure 2 shows this situation more graphically, because distributions under null and alternate hypothesis share much of their support. In any case, for individual purposes, testing is always advisable, as it may happen that the individuals are in fact genetically distant and P-values result conclusive.

Figure 3 shows the plots for a real data set from an Irish-Setter dog breed population. A total of 18 molecular markers (microsatellites) were analysed, with the number of alleles ranging from 3 to 14, the mean number being 6.94 and a mean expected heterozygosity value of 0.64. Table 6 shows the figures for power in this situation. Reasonable power is obtained only for full-sibship vs. independency comparisons for $\alpha = 0.001$.

Simulations were also performed to determine whether sampling variation on the allele frequencies affect power calculations. First, 1000 sets of random allele frequencies for 15 markers, each of them with seven alleles, were generated, and for each set, the power for the independent vs. full-sibs test was calculated. The mean power among the 1000 sets was of 0.864 ± 0.010 for $\alpha = 0.001$. This shows that very different sets of allele frequencies provide similar power estimates, so the power of the test relies in the...
A different alleles for a certain marker, say $A_1A_2$ and $A_3A_4$. It is clear that the likelihood of the two genotypes under the non-relationship hypothesis must be $2p_1p_2 \times 2p_3p_4 = 4p_1p_2p_3p_4$, where $p_i$ stands for the frequency of $A_i$, $i = 1, …, 4$. According to Goodnight and Queller’s procedure and the formulae in Table 2 of Goodnight & Queller (1999), the likelihood would equal $\frac{1}{4} p_1p_2p_3p_4 = p_1p_2p_3p_4$, which is clearly incorrect. Probably the inaccuracy lies in the allelic approach that they followed in the deduction of their formulae. In any case, their results are not comparable with ours, because they calculate the number of loci needed for a power of 0.5 and a significance level of 0.05, which is too high a figure for the usual forensic applications. Furthermore, they consider loci with 20 equally frequent alleles per locus, which is a certainly unrealistic situation to simulate.

**Conclusions**

Formulae are given to compute the probabilities ratio for different hypotheses when sibs or half-sibs are implied and parents are not known. Results, obtained under the Hardy–Weinberg assumptions and assuming that population allele frequencies are known without error, show that the amount of information generally used by the service laboratories can be sufficient to test full-sib or, with some less certainty, half-sib parentage, as more marker information will be required to reach equivalent power.

**Acknowledgements**

This work received the financial support of the EC DGVI QLRT-99-30147. Blood samples from the Irish-Setter breed were provided by the Federación Española de Caza.

**References**


