The Estimation of Parental Genotypes by the Analysis of a Fixed Number of Their Offspring

A. Barbadilla and H. Naveira

Departamento de Genética y Microbiología, Facultad de Ciencias, Universidad Autónoma de Barcelona, Bellaterra (Barcelona), Spain

Manuscript received June 30, 1987
Revised copy accepted February 24, 1988

ABSTRACT

When genotypes are cryptic in adult organisms but visible in young life-stages, it is possible to estimate the genotypes in parents after the examination of a fixed number of their offspring. Thus, in Drosophila, males collected in the wild can be crossed to females of a homozygous laboratory stock, and the identification for genotype of seven larvae is largely enough to guarantee a correct diagnosis of the two alleles of the father \((P = 0.984)\). On the other hand, when collected females were inseminated in nature, it should be possible to determine the constitution both the female and her mate, after the analysis of a sufficient number of their offspring. Nevertheless, exact formulas for the probabilities of correct diagnoses of parental genotypes in this case have not yet been published. In this paper we derive such formulas, and illustrate their application. The conclusion we draw from our calculations is that, unless 15 larvae are examined among the offspring, the probability of misdiagnosis is over the 0.05 level. Of course, the errors produced become greater as smaller offspring numbers are analyzed. But in the case of one locus with two alleles, we have derived formulas to correct the observed number of matings and estimate its actual number in the sample, for any offspring sample size. After these corrections, we have been able to calculate estimates of relative mating frequencies in the sampled population, together with their associated errors, which can be subsequently used in studies of adult fitness components, mainly in those dealing with sexual selection.

Genotypes are not always accessible to direct examination in all the stages of the life cycle of an organism. In Drosophila, for example, chromosomal polymorphism is observable only in third instar larvae. But there are many other examples of missing parental data in other organisms, usually when dealing with allozyme loci (see, for example, Luykx 1981; Samollow, Dawson and Riddle, 1983). Therefore, when we are interested in gene or genotype frequencies of the adult population, a serious problem of estimation arises. That is the case of studies on adult fitness components, like longevity (Ruiz et al. 1986), or sexual selection (Anderson and McGuire 1978). The general procedure to overcome the difficulty of cryptic genotypes in adults has consisted of collecting a sample of them, usually males, and mating them in the laboratory with flies of known homozygous genotype (Dobzhansky 1944; Anderson et al. 1975; Ruiz et al. 1986). The inference rests on the genotypes observed in the family of each collected male. Thus, about seven larvae should be observed in order to determine the two gene arrangements of the wild male (Dobzhansky 1944). Arnold (1981) has investigated the properties of two different estimators of allele probability in adults after this protocol. Later on, Arnold and Morrison (1985) described a model for another experimental protocol, which consisted of collecting \(N\) singly inseminated mothers, from each of whom \(n\) offspring were identified for genotype. Their model, which assumed the Hardy-Weinberg law, Mendelian laws, and a singly randomly chosen mate per female, provided several estimators of gene frequency in the cryptic parental gene pool.

Therefore, estimating gene frequencies in the adult mating population from family profiles is at present no problem for population geneticists (Sobel, Arnold and Sobel 1986). As a matter of fact, by simply examining the chromosomes of a single larva among the offspring of each inseminated female collected in nature, two gene arrangements, one of maternal and one of paternal origin, can be determined (Anderson et al. 1975; Dobzhansky 1958; Fontdevila et al. 1982). But there still remains the problem of how to estimate adult mating genotype frequencies. According to Dobzhansky (1944), if we examine the chromosomes of eight larvae in the offspring of a wild inseminated female, "we have a satisfactory chance" (not specified) of determining the constitution both of the female and of her mate. Anderson and McGuire (1978) examine ten larvae from each female, and although they acknowledge the possibil-
ity of misdiagnosing the genotypes of the mating couple, they claim that the probability of such event is extremely small, of the order of 0.005 (but they do not show the calculations they performed to arrive at that conclusion). In this paper we show that both ANDERSON and MCGUIRE’s and, of course, Dobzhansky’s offspring sample sizes are too small to avoid a considerable amount of misidentifications of mating genotypes.

With regard to ARNOLD and MORRISON (1985), we remove the assumption of the Hardy-Weinberg law, thus giving entry to possible deviations from random mating and to sexual selection among adults, but we keep the other two restrictions: we assume no gametic drive and we assume singly inseminated females. Our aim is to estimate adult mating frequencies in a wild population by collecting a sample of N inseminated females and analyzing in the laboratory the genotypes of a fixed number of their offspring. The table of matings thus obtained, although it is not possible to determine which karotype is male and which is female, can be tested for agreement to random mating expectations (under the assumption of equal genotype frequencies for both sexes). In addition, if an independent sample of collected adults is identified for genotype, sexual selection can be assessed.

RESULTS

Minimum number of larvae per female: Let us consider a locus with four alleles in a diploid species (four is the maximum number of different genomes implicated in a single mating), for example, a Drosophila population polymorphic for four inversions on a given chromosome (A, B, C and D). Wild inseminated females are taken to the laboratory and transferred to individual vials with standard Drosophila medium. When offspring larvae reach their third instar, a sample of n of them is dissected and the polytene chromosomes of their salivary glands are analyzed. This sample is used to infer the genotypes of the parents. The problem is to determine how many larvae should be analyzed in order to be at least 95% sure of having correctly identified the parental genotypes. Table 1 shows the different types of matings we should expect. Two broad classes can be distinguished: those where the simple presence of certain offspring genotypes (“definitive combination” of genotypes) is enough to determine unequivocally from family profiles, that is, the ratio of progeny the parental genotypes (mating types 1, 2, and 3), and those whose offspring do not offer any genotype marker of parentage (mating types 4, 5, 6, and 7), the only information on parentage being obtained from family profiles, that is, the ratio of progeny genotypes.

The offspring of each collected female are ascribed to a given mating type according to the procedure indicated in Table 2 (first and second columns). This procedure is a simple extension to four alleles of the “Dobzhansky scoring” procedure, cited by ARNOLD and MORRISON (1985).

Table 2 is organized so that the two types of mistakes we must confront can be made apparent. First, errors of type I consist of rejecting the null hypothesis when it is indeed true. In our case, that means rejecting, according to the offspring genotypes, a given mating when it did actually happen. The conditional probabilities of these errors, for each mating type, are given in the third column of Table 2 as a function of the number of larvae analyzed per collected mother.

Second, errors of type II consist of accepting the null hypothesis when it is indeed false. In our case, that means ascribing an offspring to a given mating when another one actually took place. Only the largest of the conditional probabilities of these errors, for each mating type, is shown in the fifth column of Table 2. As is seen there, this kind of error is possible

### Table 1: Different kinds of matings between genotypes for one locus with four alleles

<table>
<thead>
<tr>
<th>Mating type</th>
<th>Description</th>
<th>Example</th>
<th>Offspring</th>
<th>Definitive combinations in offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Heterozygotes sharing one allele</td>
<td>AB × AC</td>
<td>1/4 AA : 1/4 AC ; 1/4 AB : 1/4 BC</td>
<td>AA + BC ; AA + AB + AC</td>
</tr>
<tr>
<td>2</td>
<td>Identical heterozygotes</td>
<td>AB × AB</td>
<td>1/4 AA : 1/4 BB ; 1/2 AB</td>
<td>AA + BB</td>
</tr>
<tr>
<td>3</td>
<td>Heterozygotes not sharing any allele</td>
<td>AB × CD</td>
<td>1/4 AC : 1/4 AD ; 1/4 BC : 1/4 BD</td>
<td>AC + AD + BC ; AC + AD + BD</td>
</tr>
<tr>
<td>4</td>
<td>Homozygote and heterozygote sharing one allele</td>
<td>AA × AB</td>
<td>1/2 AA : 1/2 AB</td>
<td>None</td>
</tr>
<tr>
<td>5</td>
<td>Homozygote and heterozygote not sharing any allele</td>
<td>AA × BC</td>
<td>1/2 AB : 1/2 AC</td>
<td>None</td>
</tr>
<tr>
<td>6</td>
<td>Identical homozygotes</td>
<td>AA × AA</td>
<td>1 AA</td>
<td>None</td>
</tr>
<tr>
<td>7</td>
<td>Different homozygotes</td>
<td>AA × BB</td>
<td>1 AB</td>
<td>None</td>
</tr>
</tbody>
</table>
TABLE 2
Summary of results

<table>
<thead>
<tr>
<th>Offspring genotypes</th>
<th>Mating score (mating type)</th>
<th>Probability of making type I errors</th>
<th>No. of larvae to detect at least 95% of each true mating type</th>
<th>Probability of making type II errors</th>
<th>No. of larvae to avoid 95% of misdiagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least AA and BC</td>
<td>AB × AC</td>
<td>( \left( \frac{3}{4} \right)^n - 3 \left( \frac{1}{4} \right)^n + \left( \frac{1}{2} \right)^n )</td>
<td>11</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>At least AA</td>
<td>AB × AB</td>
<td>( 2 \left( \frac{3}{4} \right)^n - \left( \frac{1}{2} \right)^n )</td>
<td>13</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>At least AC, AD, and BC</td>
<td>AB × CD</td>
<td>( \frac{5}{3} \left( \frac{1}{4} \right)^{n-2} - 2 \left( \frac{1}{2} \right)^n )</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>None except AA and AC</td>
<td>AA × AB</td>
<td>( \left( \frac{1}{2} \right)^n )</td>
<td>6</td>
<td>( \left( \frac{3}{4} \right)^n - \left( \frac{1}{4} \right)^n - \left( \frac{1}{2} \right)^n )</td>
<td>11</td>
</tr>
<tr>
<td>None except AB and AC</td>
<td>AA × BC</td>
<td>( \left( \frac{1}{2} \right)^n )</td>
<td>6</td>
<td>( \left( \frac{3}{4} \right)^n - 2 \left( \frac{1}{4} \right)^n )</td>
<td>5</td>
</tr>
<tr>
<td>None except AA</td>
<td>AA × AA</td>
<td>0</td>
<td>1</td>
<td>( \left( \frac{1}{2} \right)^n )</td>
<td>5</td>
</tr>
<tr>
<td>None except AB</td>
<td>AA × BB</td>
<td>0</td>
<td>1</td>
<td>( \left( \frac{1}{2} \right)^n )</td>
<td>5</td>
</tr>
</tbody>
</table>

Underlined numbers correspond to limiting offspring sample sizes.

only when there are no genotypic markers of parentage, as happens with mating types 4, 5, 6 and 7. Taking both kinds of error into consideration, limiting offspring sample sizes can be calculated for each mating type (fourth and sixth columns, Table 2, underlined numbers).

We explain mating type 1 and type 4 in detail, and then simply refer to Table 2 for the other mating types.

Mating type 1 includes all the matings between heterozygotes (heterokaryotypes) sharing one allele (inversion), as for example AB × AC. Although four different genotypes should appear in their progeny, each with expected frequency 1 in 4 (Table 1), it is not necessary to find them all in a sample of third instar larvae in order to determine the correct parental genotypes. There are two "definitive combinations" in the offspring (Table 1): AA + BC, and AA + AB + AC. Either of these can be produced only by mating AB × AC; there is no other possibility. Therefore, our problem reduces to that of calculating how many larvae should be examined in order to be 95% sure of finding any of the two definitive combinations given that this mating had actually taken place.

This calculation is most easily performed by tree diagram analysis. Given a sample of \( n \) larvae, the probability of combination AA + BC (that is, the probability of finding at least one individual of each of these genotypes in the sample) is

\[
P(\text{AA + BC}) = 1 - \left[ 2 \left( \frac{3}{4} \right)^n + \left( \frac{1}{2} \right)^n \right] \tag{1}
\]

Similarly, for the other definitive combination (AA + AB + AC),

\[
P(\text{AA + AB + AC}) = 1 - \left[ \sum_{k=0}^{k=n-2} (\frac{3}{4})^k \frac{1}{2}^{n-2-k} \right] \tag{2}
\]

where

\[
f^{n-2-k} = \frac{1}{4} \sum_{k=0}^{k=n-2} \left( \frac{3}{4} \right)^k \left( \frac{1}{2} \right)^{n-2-k},
\]

for \( (n - 2 - k) \in \mathbb{N}_0 \)

and

\[
f^0 = 1.
\]

The joint probability of the two definitive combinations is

\[
P[\text{AA + BC} \cup \text{AA + AB + AC}] = 1 - [\left( \frac{3}{4} \right)^n - 3 \left( \frac{1}{4} \right)^n + \left( \frac{1}{2} \right)^n - 1]. \tag{3}
\]

This formula expresses the probability of arriving at the correct identification of the parental genotypes, after the analysis of \( n \) larvae, when the mating was actually of type 1. The expression in brackets corre-
sponds to the probability of making type I errors, which is given in the third column of Table 2. Also in Table 2, in the fourth column, it is shown that at least 11 larvae should be examined from each female in order to detect 95% of all these matings. On the other hand, there is no possibility of making type II errors, insofar as definitive combinations in the offspring lead to unambiguous determination of parental genotypes.

Mating type 4, as for example AA × AB, corresponds to matings between a homozygote for one allele and a heterozygote for the same allele. These same two genotypes are expected in the offspring sample, each with the same relative frequency (Table 1). Finding these two genotypes in the offspring sample is highly probable, even for relatively small sample sizes ($P = 1 - (\frac{1}{2})^{-1}$; for $n = 6$, $P = 0.969$), but this is not enough to exclude other possible parental genotypes. Table 1 shows that genotypes AA + AB are also expected after two other matings. Thus, we might have missed BB in the $n$ offspring sample after mating AB × AB (mating type 2), or both AC and BC after mating AB × AC (mating type 1). The probability of the first mistake is $(\frac{1}{2})^{n} - (\frac{1}{2})^{n} - (\frac{1}{2})^{n}$, whereas that of the second is only $(\frac{1}{2})^{n} - 2(\frac{1}{2})^{n}$, that is, approximately $(1.5)^{n}$ times smaller than the former. Mating type 4 can be confused much more easily with mating type 2 than with mating type 1. Table 2 shows that only after having examined 11 larvae can we feel reasonably confident ($P = 0.96$) of having excluded the possibility of this most important mistake.

In a similar way, limiting offspring sample sizes can be found for the other mating types. As shown in Table 2, the most disadvantageous situation, in terms of experimental effort, is the mating between two identical heterozygotes: at least 13 larvae should be analyzed in order to detect 95% of actual matings of this type, which, on the other hand, is free of type I errors, because it is always unequivocally identified. Thirteen larvae is, therefore, as a first approximation, the limiting offspring sample size in any experiment where parental genotypes are to be inferred from offspring genotypes. Observed mating frequencies in the sample can then be used directly to estimate relative mating frequencies in the population.

**Estimation of mating frequencies for a locus with two alleles:** The results just described are equally applicable to loci with more than four alleles, insofar as no novel mating type would be added to the model. When only two alleles ($A$ and $a$) for the locus in question exist in the population, the limiting mating type ($Aa \times Aa$) can be more frequent than ever. However, in this case it is not necessary to analyze 13 larvae per female to get rather good estimates of mating frequencies. The importance of the different sources of error in the assignment of parental genotypes to each family in this case is easy to quantify as a function of the number of larvae analyzed per female. Corrections can be performed to calculate estimated mating frequencies from the observed frequencies in the sample.

We refer now more closely to the model of ARNOLD and MORRISON (1985). We have two alleles, $A$ and $a$. A collected mother carries $y$ copies of the $A$ allele. The count $y$ must be 0, 1, or 2, and it can be identified with the cryptic genotype of the collected female. We assume that each mother mates with only one father in nature. The parental genotypes are denoted by $y_0$ and $y_1$. These parental genotypes $y = (y_0, y_1)$ are cryptic to the collector, and it is not possible to distinguish which is the father and which is the mother.

For a given mother, examination of $n$ of her offspring yields a profile $\pi = (n_0, n_1, n_2)$ where (i) $n$ offspring are homozygous for the $A$ allele; (ii) $n$ are heterozygous; (iii) $n$ are homozygous for the $A$ allele. The family profiles $\Lambda = \{(n_0, n_1, n_2) : n_0 \geq 0, n_1 \geq 0, n_2 \geq 0, n_0 + n_1 + n_2 = n\}$ are defined in a sample space $\Lambda$. Given a family profile $\pi$, it is possible to select parental genotypes $y$ so that the conditional density $\text{prob}(\pi | y)$ is maximum. When the mother mates only once and the Mendelian laws of inheritance are valid, the procedure to maximize $\text{prob}(\pi | y)$ is called the "Dobzhansky scoring procedure," and the score equals $y_0 \times y_1$. This procedure is shown in Table 3.

Six mutually exclusive events $\Lambda_{00}, \ldots, \Lambda_{11}$ are singled out for special consideration (Table 3). In this table, $\Theta_{00}, \ldots, \Theta_{11}$ are the probabilities of the different kinds of mating in the wild population; $\alpha_1 = 2(\frac{1}{2})^n - (\frac{1}{2})^n$; $\alpha_2 = (\frac{1}{2})^{n-1}$; and $\alpha_3 = (\frac{1}{2})^n - (\frac{1}{2})^n - (\frac{1}{2})^n$, where $\alpha_i$ are the probabilities of making wrong determinations. The events $\Lambda_{00}, \ldots, \Lambda_{11}$ constitute a partition of the different possible family profiles $\pi$ in the sample space $\Lambda$. This partition identifies family profiles by the presence or absence of one or more genotypes in a family. The probabilities $K_{00}, \ldots, K_{11}$, of each of these six events can be computed from the model specification

$$\text{prob}(\pi | \Theta_{00}, \ldots, \Theta_{11}) = \text{prob}(\pi | y) \text{prob}(y | \Theta_{00}, \ldots, \Theta_{11}),$$

and are shown in the second column of Table 3. In a collection of $N$ mothers, the numbers $N_{00}, \ldots, N_{11}$ count types of family profiles based on the presence or absence of one or more genotypes. The probability density of $N = (N_{00}, \ldots, N_{11})$ is multinomial, and is given by

$$\text{prob}(N | \Theta_{00}, \ldots, \Theta_{11}) = \left(\begin{array}{c} N \\ N_{00} \ N_{02} \ K_{22}^N \ K_{01}^N \ K_{11}^N \end{array} \right)$$

where $\frac{N_{00}}{N} = \frac{N!}{N_{00}! N_{02}! N_{22}! N_{01}! N_{11}!}$.
### TABLE 3

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
<th>Count (N)</th>
<th>Score</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{00} = {n: n_0 = n}$</td>
<td>$K_{00} = \theta_{00} + \theta_{11}(\frac{1}{4})^n + \theta_{01}(\frac{1}{2})^n$</td>
<td>$N_{00}$</td>
<td>$0 \times 0$</td>
<td>AA × AA</td>
</tr>
<tr>
<td>$A_{12} = {n: n_1 = n}$</td>
<td>$K_{02} = \theta_{02} + (\theta_{11} + \theta_{01})$</td>
<td>$N_{02}$</td>
<td>$0 \times 2$</td>
<td>AA × aa</td>
</tr>
<tr>
<td>$A_{22} = {n: n_2 = n}$</td>
<td>$K_{22} = \theta_{22} + \theta_{01}(\frac{1}{4})^n + \theta_{10}(\frac{1}{4})^n$</td>
<td>$N_{22}$</td>
<td>$2 \times 2$</td>
<td>aa × aa</td>
</tr>
<tr>
<td>$A_{01} = {n: n_0 + n_1 = n, n_0 &gt; 0, n_1 &gt; 0}$</td>
<td>$K_{01} = \theta_{00}(1 - a_2) + \theta_{11}a_3$</td>
<td>$N_{01}$</td>
<td>$0 \times 1$</td>
<td>AA × Aa</td>
</tr>
<tr>
<td>$A_{12} = {n: n_1 + n_2 = n, n_1 &gt; 0, n_2 &gt; 0}$</td>
<td>$K_{12} = \theta_{02}(1 - a_2) + \theta_{11}a_3$</td>
<td>$N_{12}$</td>
<td>$1 \times 2$</td>
<td>Aa × aa</td>
</tr>
<tr>
<td>$A_{11} = {n: n_0 + n_1 + n_2 = n, n_0 &gt; 0, n_1 &gt; 0, n_2 &gt; 0}$</td>
<td>$K_{11} = \theta_{11}(1 - a_1)$</td>
<td>$N_{11}$</td>
<td>$1 \times 1$</td>
<td>Aa × Aa</td>
</tr>
</tbody>
</table>

$\theta_{00}, \theta_{01}, \theta_{02}, \theta_{11}, \theta_{12}, \theta_{22}$: probabilities of different matings. For one locus and two alleles, $2 = \text{homozygote for one allele}, 0 = \text{homozygote for the other allele}, 1 = \text{heterozygote}$. $a_1 = 2(\frac{1}{4})^n - (\frac{1}{2})^n; a_2 = (\frac{1}{2})^{n-1}; a_3 = (\frac{1}{4})^n - (\frac{1}{2})^n - (\frac{1}{4})^n$.

For a given experimental effort, $K(K = N \cdot n; N = \text{number of collected females}; n = \text{number of larvae analyzed per female})$, it is possible to find the optimum distribution of effort between $N$ and $n$, that is, the values $n$ and $N = K/n$ which produce a minimum variance of the estimator of a given mating probability. These values always depend on the corresponding parameter. For example, as is shown in Figure 1, the values of $n$ and $K/n$ which lead to a minimum variance of the estimator of the probability of matings between two heterozygotes ($\text{Var}(\hat{\theta}_{11})$), depend on $\theta_{11}$, the corresponding parameter. As it is shown there, when $\theta_{11}$ is high, the best strategy is to collect fewer females and analyze more larvae from each female. Similar figures can be obtained for the other mating types, sometimes with opposite results (A. Barbadilla and H. Naveira, in preparation).

### AN ILLUSTRATION

DOBZHANSKY et al. (1963) examined six larvae per progeny of wild inseminated females of *D. pseudoobscura* from several sampling stations in Bogotá (Colombia). Their pooled results are shown in Table 4. From these results (observed frequencies, with no kind of correction), the authors concluded that adults did not mate at random. This conclusion was, nevertheless, complicated by the observation that putative genotype frequencies of the adult population were extremely similar to the frequencies expected under Hardy-Weinberg equilibrium, which implies random mating. Our analysis shows that both these two results may be spurious. Our estimates of the different mating frequencies are given in Table 4, under the observed frequencies, together with their corresponding standard errors. It is immediately apparent that there is a large difference between observed and estimated numbers of the SC/TL × SC/TL mating. Of course, relative frequencies of the two gene arrangements do not change after our corrections, although genotype frequencies do. These corrected genotype frequencies are significantly different from...
those expected under Hardy-Weinberg ($\chi^2 = 7.874$, 1 d.f., $P = 0.005$), due to an excess of heterozygotes over expected. If genotype frequencies are equal for both sexes, genotypes still appear not to be mating at random ($\chi^2 = 9.368$, 3 d.f., $P = 0.025$). On the one hand, deviation of genotype frequencies from Hardy-Weinberg expectations in the parental population may be due to sexual selection acting on males, which Anderson et al. (1979) have shown to be an important evolutionary force in this species, although it could be also explained by heterosis in viability (Dobzhansky and Levene 1948). This question might have been easily resolved by taking an additional sample of genotypes in the adult male population.

On the other hand, the significant deviation from what we have called, following Dobzhansky et al. (1963), “random mating frequencies,” may be spurious. Just as Dobzhansky et al. did, we have derived expected mating frequencies under the assumption that genotype frequencies are the same for both sexes. There is no alternative, because with our technique for inferring the parental genotypes we cannot discern which genotype is the father and which is the mother. The lack of fit shown by the $\chi^2$ test might easily be due to different genotype frequencies between males and females in the mating couples. If there is indeed sexual selection among males, our test might produce significant deviations from expected mating frequencies, even if males were mating effectively at random with females. This is our contention. The results of Dobzhansky et al. (1963) do not prove conclusively that adults did not mate at random; most probably, they reflect only different genotype frequencies between males and females engaged in mating, differences produced by sexual selection among the males. Again, this question could have been resolved only by taking an additional sample of genotypes from the adult male population.

**DISCUSSION**

We give exact formulas for the probabilities of the different kinds of errors associated with the estimation of adult mating frequencies from observed offspring genotypes. For a single locus with two alleles it is possible to estimate actual mating frequencies with arbitrary small standard errors, by choosing the offspring sample size, the number of observed matings, or both these numbers.

When more than two alleles for the locus in question are segregating in the population, the offspring sample size per collected female should not be smaller than 13, unless the corrections derived for the two allele model were also applied here. Then 11 offspring genotypes should be examined in order to minimize the probability of type I errors associated with mating type 1 (Table 2).

The derivations of our formulae rest on several assumptions, which may be summarized as follows:

- No gametic drive.
- No viability differences among karyotypes in the laboratory.
- No multiple insemination of wild-caught females, or, at least, strong predominance among the offspring of the progeny sired by the last mated male.

By saying this, we do not exclude the possibility of extending the range of application of these methods of estimation. Thus, for example, deviations from Mendelian segregation laws can be easily incorporated into our model. If the magnitude of the drive was known, it would be very simple to calculate a new set of formulas for the expected errors of diagnosis of each mating type. Something similar can be said about viability differences among karyotypes. The problem of multiple inseminations is more difficult to resolve. Fortunately, sperm displacement, or whatever it is the mechanism of sperm predominance, seems to be a well established fact in Drosophila and other organisms (Smith 1984), and so it happens that when a wild-caught female is placed in a culture bottle, only rather seldom is the batch of eggs deposited by her fertilized by the sperm of more than one father (Dobzhansky et al. 1963).
If we desired to estimate the genotypes for several loci of each mating pair, formulas might be derived by applying essentially the same rationale we applied here for one locus, although the complication would be considerably increased by the much higher number of different offspring genotypes and mating types.

We wish to thank B. S. Weir and two referees for their comments on this manuscript. This material is based on work supported by grant PB85-0071 from the Comisión Interministerial de Ciencia y Tecnología (CICYT), Spain to A.F.

**LITERATURE CITED**


Communicating editor: B. S. Weir

**APPENDIX**

For the case of one locus and two alleles the variances associated to estimators of the different mating type probabilities are the following:

\[
\text{Var}(\hat{\theta}_{11}) = \frac{(1 - \alpha_1)^2 \text{var}(a) + \text{var}(b)}{N^2(1 - \alpha_1)^2}
\]

\[
\text{Var}(\hat{\theta}_{01}) = \frac{(1 - \alpha_2)^2 \text{var}(c) + \text{var}(d)}{N^2(1 - \alpha_2)^2} + \frac{\text{var}(e) + \left[\alpha_2^2(1 - \alpha_1)^2\right] \text{var}(b)}{N^2(1 - \alpha_2)^2}
\]

\[
\text{Var}(\hat{\theta}_{12}) = \frac{(1 - \alpha_2)^2 \text{var}(i) + \text{var}(j) + \text{var}(e)}{N^2(1 - \alpha_2)^2} + \frac{\left[\alpha_2^2(1 - \alpha_1)^2\right] \text{var}(b)}{N^2(1 - \alpha_2)^2}
\]

\[
\text{Var}(\hat{\theta}_{00}) = \frac{1}{N^2} \left[ \text{var}(f) + \text{var}(g) + \text{var}(h) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(d) + \text{var}(e) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(b) \right]
\]

\[
\text{Var}(\hat{\theta}_{22}) = \frac{1}{N^2} \left[ \text{var}(k) + \text{var}(m) + \text{var}(h) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(j) + \text{var}(e) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(b) \right]
\]

\[
\text{Var}(\hat{\theta}_{02}) = \frac{1}{N^2} \left[ \text{var}(p) + \text{var}(q) + \text{var}(g) + \text{var}(m) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(b) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(d) \right.
\]

\[
+ \left[ \alpha_2^2 \frac{\alpha_2^2}{1 - \alpha_1}\right] \text{var}(b) + 2 \text{var}(e) + \text{var}(j) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(b) + \left(\frac{\alpha_2^2}{1 - \alpha_1}\right) \text{var}(b)
\]

where:

\[
\text{Var}(a) = N\Theta_{11}(1 - \Theta_{11})
\]

\[
\text{Var}(b) = N\Theta_{11}\alpha_1(1 - \alpha_1)
\]
Two kinds of error are incorporated in all these variances: that due to the offspring sample size per collected female (error of diagnosis of parental genotypes) and that due to the size of the sample of inseminated females (error of sampling from the population of inseminated females). Both kinds of error affect the estimation of mating frequencies in the population.

As an example of the derivation of these formulas we examine

\[ \text{Var}(\hat{\Theta}_{11}) = \frac{(1 - \alpha_1)^2 \text{Var}(a) + \text{Var}(b)}{N^2(1 - \alpha_1)^2} \]

which corresponds to the matings between two identical heterozygotes, for example \( Aa \times Aa \).

Since \( \hat{\Theta}_{11} = N_{11}/N(1 - \alpha_1) \), we have

\[ \text{Var}(\hat{\Theta}_{11}) = \text{Var}[N_{11}/N(1 - \alpha_1)] = \text{Var}[k_{11}/(1 - \alpha_1)] = [\text{Var}(k_{11})]/(1 - \alpha_1)^2. \]

The variance of \( k_{11} (N_{11}/N) \) correspond to a binomial distribution, so

\[ \text{Var}(k_{11}) = [k_{11}(1 - k_{11})]/N \quad \text{and} \]
\[ \text{Var}(\hat{\Theta}_{11}) = [k_{11}(1 - k_{11})]/N(1 - \alpha_1)^2. \]

This last formula can be expressed as

\[ \text{Var}(\hat{\Theta}_{11}) = \frac{(1 - \alpha_1)^2 \text{Var}(a) + \text{Var}(b)}{N^2(1 - \alpha_1)^2} \]

where

\[ \text{Var}(a) = N\Theta_{11}(1 - \Theta_{11}) \quad \text{and} \]
\[ \text{Var}(b) = N\Theta_{11}\alpha_1(1 - \alpha_1). \]