Sampling Error, Effective Paternity, and Estimating the Genetic Structure of Honey Bee Colonies (Hymenoptera: Apidae)

DAVID R. TARPY[1] AND DAVID. I. NIELSEN[2]

Department of Neurobiology and Behavior, Cornell University, Seeley G. Mudd Hall, Ithaca, NY 14853

Ann. Entomol. Soc. Am. 95(4): 513–528 (2002)

ABSTRACT Multiple mating by social Hymenopteran queens significantly lowers the average genetic relatedness among female nestmates, which subsequently affects a wide range of social behaviors. Honey bees (*Apis* spp.) have among the highest levels of multiple mating in social insects, and have received the most empirical effort to quantify the effective paternities within colonies. We reviewed 24 studies that estimated paternity frequencies of individual, naturally-mated honey bee queens using molecular techniques. We summarize the methods used to estimate effective paternity (m_e) and intracolonial genetic relatedness (G). We then concentrate on the effect of sample size on estimates of m_e using Monte Carlo simulations. The results demonstrate that m_e estimates may vary significantly as a result of sampling error, particularly at low worker sample sizes and high paternity numbers. From these simulations, we arbitrarily define a "threshold" worker sample size to effective paternity ratio (n/m_e) that, at best, reduces the error of estimating m_e to less than one subfamily. The literature review illustrates that no study with an n/m_e ratio above this threshold estimates an m_e above 15 subfamilies. Finally, we briefly discuss other factors that may serve to over-estimate m_e, including numerous sampling biases. We conclude that although 152 colonies in the various species of *Apis* have been tested, the extremity of their paternity frequencies may be somewhat exaggerated, although not drastically.

KEY WORDS honey bee, polyandry, multiple mating, paternity analysis, mating number

KIN SELECTION HAS become the most widely accepted explanation of social evolution among insects (Hamilton 1972, West-Eberhard 1975, Bourke and Franks 1995, Crozier and Pamilo 1996), but the hypothesis is not without its caveats. One issue that has confronted kinship theory is the existence of polyandry among the social Hymenoptera (Page 1986, Strassmann 2001). Multiple matings by queens reduce the high degrees of genetic relatedness that exist between haplodiploid female nestmates and significantly decrease the inclusive fitness of workers (reviewed by Crozier and Pamilo 1996).

Although low levels of polyandry occur across the social Hymenoptera, ubiquitously high mating frequencies of queens are generally isolated to the highly eusocial genera *Apis*, *Vespula*, *Acromyrmex*, and *Atta* (Boomsma and Ratnieks 1996, Boomsma et al. 1999). Honey bees (*Apis* spp.) have long been the most studied of these groups, where initial techniques to estimate mating number include direct observation (Roberts 1944, Gary 1963), sperm counts (Woyke 1962, Koeniger et al. 1990, Koeniger et al. 1994), and variant population genetics (Taber 1954, Taber and Wendel 1958, Adams et al. 1977). However, these techniques can be subject to significant imprecision. For example, variance estimates using phenotypic markers of populations are sensitive to deviations from Hardy-Weinberg equilibrium, and are therefore most likely to vary greatly (Adams et al. 1977).

Within the last decade, modern molecular techniques have enabled researchers to estimate more accurately the mating number of individual queen bees (reviewed in Oldroyd et al. 1998, Haberl and Tautz 1999). A queen's mating number is typically inferred by determining the genotypes of her worker offspring and tabulating the different paternal marker sets. Different techniques and procedures have been used among studies and collectively have produced a wide range of observed mating numbers within and among *Apis* species (Fig. 1).

Although 'mating number' is a generic and somewhat ambiguous term (see Boomsma and Ratnieks 1996), the important issue—at least to genetic relatedness and social behavior—is the effective paternity frequency of queens. Effective paternity frequency, m_e, is the number of mates by a queen if all of her mates are equally represented in her offspring (see derivations below). A queen that is inseminated with five males could have drastically different effective paternity frequencies if each male sires an equal percentage

[1] E-mail: dt66@cornell.edu.
[2] Department of Entomology, University of California, Davis, CA 95616.

0013-8746/02/0513–0528$02.00/0

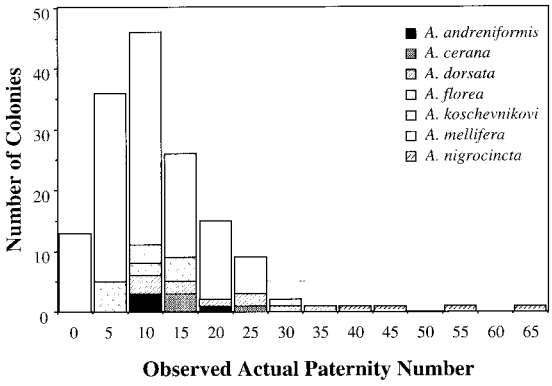


Fig. 1. Observed actual paternity numbers of individual colonies of seven species in the genus *Apis*. All estimates are based on studies that used molecular techniques to determine the genotypes of worker offspring from individual, naturally-mated queens (reviewed in Appendices).

of her offspring ($m_e = 5.00$) or if one male sires 99% of her offspring ($m_e \approx 1.02$).

Although modern molecular techniques are more precise than previous methods, they are also subject to some uncertainty. Specifically, experimental sampling of worker offspring from colonies leads to sampling error (where subfamilies are not represented in their true proportions) and nondetection error (where the marker type of one or more subfamilies are identical to one another). These uncertainties have led several researchers to conclude that the current estimates of effective paternity frequencies in honey bees are conservative (e.g., Boomsma and Ratnieks 1996, Oldroyd et al. 1998).

Little attention, however, has been given to factors that may potentially inflate estimates of effective paternity. There is some anecdotal evidence that this has occurred, because some estimates exceed the limits of honey bee reproductive biology. For example, an estimated effective paternity of 40 in an *Apis mellifera* (L.) colony (see *Appendix* 1) would require that the queen obtains equal sperm loads from 40 males. An average male's sperm load is ≈1.0 μl, and a queen's temporary storage capacity for semen in her lateral oviducts is about 10 μl on a single mating flight. Assuming that sperm enter into and are sampled from the spermatheca equally (Laidlaw and Page 1984), a queen would have to completely fill up with sperm on at least four mating flights, a phenomenon that is exceedingly rare (Roberts 1944, Woyke 1964).

The purpose of this article is to highlight how certain factors may serve to overestimate effective paternity frequencies in social insect colonies by using honey bees as a focal taxon, both because of their high levels of polyandry and the large number of studies quantifying effective paternity in the genus *Apis*. We review the techniques that researchers have used to estimate effective paternity frequency and intracolonial genetic relatedness in such studies. We then focus on how sampling error affects estimates of effective paternity by using Monte Carlo simulations. Finally, we briefly discuss other factors that may impact estimates of effective paternity frequency. This process will determine the extent of our current understanding of polyandry among honey bee species as well as provide a general framework from which molecular studies of insect colonies should estimate effective paternity and genetic relatedness.

Table 1. Descriptions of variables used to estimate effective paternity and intracolonial genetic relatedness

Variable	Description
n	Number of worker samples
N	True actual paternity number
N_o	Observed actual paternity number
$\hat{N}_e$	Estimated actual paternity number (equation 7)
m_e	True effective paternity frequency
$\hat{m}_{e(s)}$	Estimated effective paternity frequency by equation 1
$\hat{m}_{e(p)}$	Estimated effective paternity frequency by equation 2
$\hat{G}$	Estimated average genetic relatedness among female nestmates
$\hat{m}_{e(s)} \rightarrow \hat{G}$	Direct method of calculating the statistic of m_e. Estimate m_e first using equation 1, then G using equation 3
$\hat{m}_{e(p)} \rightarrow \hat{G}$	Direct method of calculating the sample statistic of m_e. Estimate m_e using equation 2, then estimate G using equation 3
$\hat{G} \rightarrow \hat{m}_{e(s)}$	Indirect method of calculating the statistic of m_e. Estimate G using equation 4, then estimate m_e using equation 5
$\hat{G} \rightarrow \hat{m}_{e(p)}$	Indirect method of calculating the sample statistic of m_e. Estimate G using equation 6, then m_e using equation 5

Review of Techniques that Estimate Effective Paternity

The following is an effort to summarize the methods that have been used previously to estimate effective paternity and genetic relatedness within honey bee colonies. This is neither an attempt to review all of the possible techniques used to estimate genetic relatedness in social insects, nor is it an attempt to propose novel formulae. For more comprehensive discussions of the methods used to calculate genetic relatedness in social insect colonies, see Pamilo and Crozier (1982), Queller and Goodnight (1989), Pedersen and Boomsma (1999), and references therein.

Table 1 summarizes the notation that will be used throughout this article to discuss paternity and relatedness in honey bee colonies. The two important measures—effective paternity frequency of the offspring (m_e) and the average genetic relatedness among female nestmates (G)—are interdependent, and several approaches have been used to calculate them. Many researchers have used the statistic developed by Starr (1984) to estimate m_e, where

$$\hat{m}_{e(s)} = \frac{1}{\sum_{i=1}^{N} p_i^2}, \qquad [1]$$

where N is the number of drone fathers represented within the offspring, and p_i is the relative proportion

of offspring sired by male *i*. This statistic is always less than the actual number of mates unless there is no variation in male representation.

Like all statistics, equation 1 is subject to sampling error at finite sample sizes. This has prompted several researchers to develop sample statistics, but we consider here only the one that has been used in most studies that estimate *Apis* paternity. Pamilo (1993) stated that an unbiased sample statistic of m_e is

$$\hat{m}_{e(p)} = \frac{n\text{-}1}{n\sum_{i=1}^{N}\hat{p}_i^{\,2} - 1}, \qquad [2]$$

where n is the number of workers sampled, and p_i is the estimated relative proportion of offspring sired by male i ($\hat{p}$ in equation 2 is often denoted as y_i to distinguish from p_i in equation 1). As Pamilo (1993) and Pedersen and Boomsma (1999) point out, this sample statistic should be unbiased only for population-wide estimates of m_e but has been widely used in studies that estimate paternity of individual *Apis* queens.

The average genetic relatedness among female nestmates, G (Crozier 1970), can be calculated directly from an estimate of m_e, such that

$$\hat{G} = \frac{1}{4} + \frac{1}{2\hat{m}_e}. \qquad [3]$$

When $m_e = 1$ (monandry), $G = 0.75$ because honey bees are haplodiploid, and thus females share all of their genes in common from their fathers. Any m_e value >2 causes G to lie below 0.5, and G approaches 0.25 as m_e approaches infinity.

Several studies have used the above approaches to estimate m_e directly and G indirectly (see *Appendices* 1 and 2). However, other studies have derived m_e estimates indirectly after estimating G directly. Laidlaw and Page (1984) showed that

$$\hat{G} = \sum_{i=1}^{N} p_i[0.75p_i + 0.25(1 - p_i)]. \qquad [4]$$

This estimate of G can then be used to estimate m_e by simply solving equation 3 for m_e, such that

$$\hat{m}_e = \frac{2}{4\hat{G} - 1} \qquad [5]$$

Since the proportions in equation 4 are assuming true paternity frequencies, this estimate of m_e is equal to $m_{e(s)}$ as calculated by equation 1, and can be easily verified computationally.

Several researchers have also estimated G directly following a procedure outlined by Estoup et al. (1994). For example, Neumann et al. (1999b) "determined the pedigree coefficients of relatedness G (Pamilo and Crozier 1982) between all possible worker dyads in the sample (either 0.25 or 0.75) and calculated the arithmetic mean to obtain the average intracolonial coefficient of relatedness, "G." We express this notationally as

$$\hat{G} = \frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} r_{ij} \qquad [6]$$

where $r_{ij} = 0.75$ if individuals i and j are in the same subfamily (super sisters), and $r_{ij} = 0.25$ if they are in different subfamilies (half sisters). Using equation 5, this produces an equivalent estimate of m_e to the sample statistic $\hat{m}_{e(s)}$ since, like equation 2, it takes into account finite worker sample sizes, and also can be verified computationally.

One final method is to estimate the actual paternity number within a colony, because some subfamilies may not be represented in a finite sample of offspring. Assuming that the fertilization frequencies of the males are equal within a brood, the expected number of observed drone fathers within a particular sample, $E(N_o)$, can be estimated by

$$E(N_o) = \hat{N}_e - \left[\hat{N}_e\left(1 - \frac{1}{\hat{N}_e}\right)^n\right] \qquad [7]$$

(Oldroyd et al. 1997, Cornuet and Aries 1980; see also Moritz et al. 1995 for a different but equivalent method). (The notation of N_e is often given as k but is purposefully different here to distinguish N_e from the number of sex alleles in a population). The estimated actual paternity number, N_e, may be determined for a given n by substituting N_o for $E(N_o)$ and performing iterative calculations of equation 7. Because males are assumed to be equally represented in the offspring, this calculation assumes that $N_e = m_e$, enabling G to be calculated using equation 5. This is a questionable approach, however, because subfamilies are not typically equal in frequency (see below). Besides, equation 7 assumes equal paternity, whereas m_e infers unequal paternity, which is teleological.

To summarize, there are two basic calculations used to estimate the effective paternity frequency of honey bee queens, either the statistic $\hat{m}_{e(s)}$ or sample statistic $\hat{m}_{e(p)}$. Furthermore, there are two approaches used to calculate each estimate of $\hat{m}_e$, either directly (using equations one or 2, respectively) or indirectly (by first estimating G with equations four or 6, respectively).

Effect of Sample Size on Estimates of Effective Paternity

Sampling Error. Experimental error may be introduced while estimating m_e if one or more subfamilies are not represented within a particular worker sample or if the subfamilies are not represented in their true proportions, which is why sample statistics such as equations two and seven have been developed. The purpose here is to explore the effect of sampling error on m_e estimates using Monte Carlo simulations.

A True Basic computer simulation was written to sample repeatedly workers from theoretical colonies. The program varied independently paternity number (N) and worker sample size (n). Workers were sampled randomly from a colony that was assumed to be infinitely large. The representation of each male was

uniform in the brood, i.e., each male had a $1/N$ chance of siring each worker. We initially define paternity to be equal for simplicity, even though uniform representation is rare among honey bee drone fathers (see Oldroyd et al. 1998). We cover the effect of unequal paternity on m_e estimates with a similar set of simulations in a later section (see below). The present simulations, therefore, are analogous to randomly sampling (with replacement) single sperms from the spermatheca of queens that mated with N equally frequent males a total of n times.

The estimated proportion of each subfamily within a sample (p_i) was calculated from the generated data, and the two estimates of effective paternity frequency, $\hat{m}_{e(s)}$ and $\hat{m}_{e(p)}$, were calculated with equations one and 2, respectively. It is necessary to note that $\hat{m}_{e(p)}$ could not be calculated if all of the sampled individuals belonged to different subfamilies (i.e., $n_1 = n_2 = \cdots = n_N = 1$). This is because the sum of squared proportions, $\Sigma \hat{p}_i^2$, equals $1/n$, which causes the denominator in equation 2 to equal zero. In other words, any simulation where all subfamilies were represented by a single individual would cause the calculation of $\hat{m}_{e(p)}$ to be infinitely large. Therefore, simulations where $\hat{m}_{e(p)} = \infty$ (typically at very low sample sizes and high paternity numbers) were repeated.

The simulation performed 1,000 iterations for N = 2, 5, 10, and 20 each with n = 2–200, and the means of $\hat{m}_{e(s)}$ and $\hat{m}_{e(p)}$ were calculated for each paternity number/sample size combination. The upper and lower 95% CL of both estimates were determined by taking the highest and lowest 25 values obtained from the 1,000 iterations. These confidence limits were asymmetrical at low sample sizes, but became symmetrical at higher sample sizes. Standard deviations were also obtained around each mean $\hat{m}_{e(p)}$ and used to generate confidence intervals in the *Appendices* (see below).

The computer simulations illustrate the effect of sampling error on m_e estimates (Fig. 2), and several results are worth noting. First, $\hat{m}_{e(s)}$ is a poor estimator of effective paternity frequency, because it underestimates m_e, particularly at small sample sizes. This is not surprising, because the statistic is only valid with infinitely large samples and justifies why $\hat{m}_{e(p)}$ has been the preferred calculation in empirical studies. Second, the confidence intervals of both m_e estimates is an increasing function of paternity number. This is also an intuitive result, because it requires a larger sample size to determine accurately a larger number of subfamilies within a colony. Third, the sample statistic of m_e is very sensitive to sampling error. Although the mean $\hat{m}_{e(p)}$ values approach the true effective paternity frequency at relatively small sample sizes, the variances around the means are quite large even at large sample sizes. Note that the 95% CI of $\hat{m}_{e(p)}$, in contrast to $\hat{m}_{e(s)}$, almost always overlap the true m_e, illustrating why $\hat{m}_{e(p)}$ is the more desirable technique to estimate m_e. However, any singular estimate of $\hat{m}_{e(p)}$ may vary significantly because of sampling error, reducing the certainty of individual estimates. Finally, the 95% confidence intervals (CI) of $\hat{m}_{e(p)}$ are asymmetrical around the means at low sample sizes. With the exception of N = 2, the majority of the measurements lie above the true m_e.

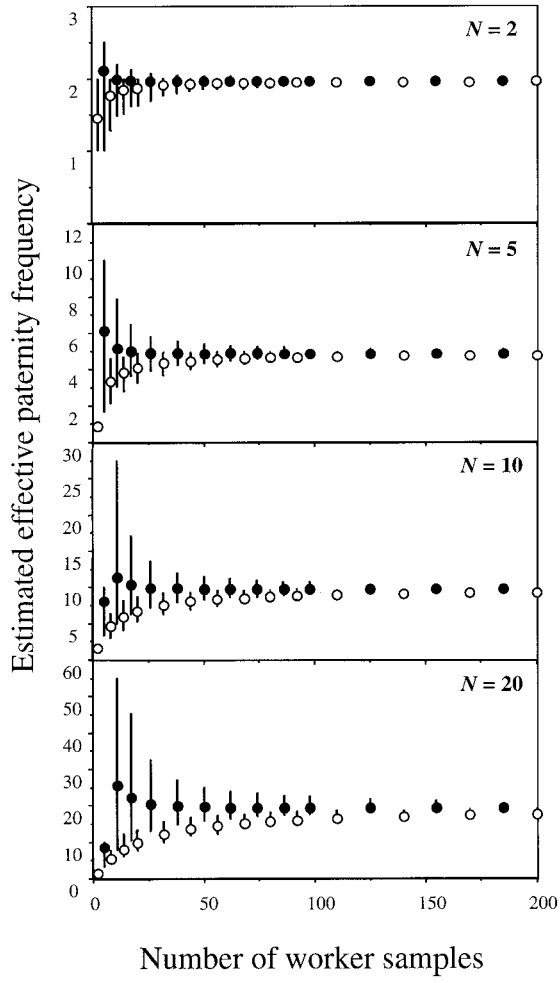


Fig. 2. Effect of sampling error on estimates of effective paternity frequency. Results are from simulations that sampled n workers from theoretical colonies with N equally-represented subfamilies. ○ = $\hat{m}_{e(s)}$; ● = $\hat{m}_{e(p)}$; both with 95% confidence intervals. Sample sizes that are illustrated: $\hat{m}_{e(s)}$, n = 2, 8, 14, 20, 32, 44, 56, 68, 80, 92, 110, 140, 170, 200; $\hat{m}_{e(p)}$, n = 5, 11, 17, 26, 38, 50, 62, 74, 86, 98, 125, 155, 185.

Indeed, some estimates of $\hat{m}_{e(p)}$ may be grossly inflated at small sample sizes. For example, it is within 95% CI that $\hat{m}_{e(p)}$ may be 3.5 times higher than the true m_e if there are 20 equally represented subfamilies in a colony and only 14 workers are sampled (not shown in Fig. 2). Certainly, an m_e estimate of 70 from 14 workers is questionable, and few studies have sampled such small numbers of workers. This error can still result in an overestimate of $\hat{m}_{e(p)}$ by 30% at a sample size of 40, illustrating the potential magnitude of overestimation of m_e at the median sample size of all studies. This is not to say that sampling error does not often underestimate m_e. Rather, underestimates of m_e

do not stand out, whereas extraordinarily large estimates of m_e have drawn great attention.

The error of $\hat{m}_{e(p)}$ decreases as sample size increases. Moreover, the magnitude of the error is a function of the true effective paternity frequency, which increases as N increases. Therefore, the ratio of n to m_e can be used to determine a standard measure of the magnitude of error in $\hat{m}_{e(p)}$. We define a "threshold" measure as the $n/\hat{m}_{e(p)}$ ratio at which the upper 95% CL is ≤1 effective drone father. In other words, this "threshold" value is the sample size to $\hat{m}_{e(p)}$ ratio where 95 of 100 iterations would produce an estimate of $\hat{m}_{e(p)}$ that does not deviate more than one subfamily from the true m_e. Using Monte Carlo simulations, this threshold was determined at equal paternity frequencies N = 3–19. The simulation performed 1,000 iterations at each value of N, incrementally increasing the sample size until the highest 25th value was no greater than one subfamily from the true N. An average threshold value from 25 independent simulations was then calculated at each paternity number. The logarithmic function

$$n/\hat{m}_{e(p)} = 9.082 \text{ Log}_{10}[\hat{m}_{e(p)}] - 1.876, \qquad [8]$$

approximates the threshold value of $n/\hat{m}_{e(p)}$ for any given m_e ($r^2 = 0.990$, $P < 0.0001$). For an estimated effective paternity of ten, the threshold value would be 7.206 workers per effective drone father. We should note that the logarithmic function asymptotes more quickly than the average threshold ratios, thus this approximation is likely to underestimate the threshold $n/\hat{m}_{e(p)}$ for effective paternity frequencies >20.

Accuracy of Previous Effective Paternity Estimates. Given the variability of $\hat{m}_{e(p)}$ as a consequence of finite sample size, it would be interesting to determine which estimates of effective paternity that have been reported for *Apis* colonies are more subject to error. To this end, 24 studies were compiled that estimated the number of subfamilies within individual colonies of *Apis* species. This data set includes only studies that used molecular techniques to determine the paternity of worker offspring from individual, naturally mated queens. Such techniques include DNA fingerprinting, restriction fragment-length polymorphism (RFLP), allozyme electrophoresis, and polymerase chain reaction (PCR)-based techniques such as RAPD-PCR and microsatellites. If possible, the effective paternity frequency of each queen was calculated directly from the reported data. If the proportional data were unavailable, the methods used to estimate m_e by the authors were verified to be equivalent to $\hat{m}_{e(p)}$ (see *Appendices*). The ratio of n to $\hat{m}_{e(p)}$ was calculated for each colony to determine how many workers per effective drone father were tested to estimate $\hat{m}_{e(p)}$. The $\hat{m}_{e(p)}$ estimates for each of 149 queens is plotted against their $n/\hat{m}_{e(p)}$ ratios in Fig. 3.

The curve in Fig. 3 is the logistic function that denotes the "threshold" $n/\hat{m}_{e(p)}$ (see above). All points above this curve are below the threshold (indicating that the errors of $\hat{m}_{e(p)}$ are likely to be greater than one subfamily), and all points on or below the curve are above the threshold (indicating that the errors of $\hat{m}_{e(p)}$ are likely to be less than one subfamily). Only 32.9% of these colonies have $n/\hat{m}_{e(p)}$ ratios that are large enough to be 95% certain that they estimate m_e to within one subfamily. It is interesting that the highest estimates of m_e above the $n/\hat{m}_{e(p)}$ threshold are 10.36 for *A. mellifera* and 14.72 for *A. dorsata* (F.). All of the higher estimates of m_e derive from studies where the $n/\hat{m}_{e(p)}$ ratio lies below the threshold. This result questions the reliability of the estimates of extreme polyandry in *Apis*. Indeed, it is possible that many of the m_e estimates above 15 are significantly overestimated as a result of sampling error. This is not to say that there are no colonies that have effective paternities that exceed this level. This result partially reflects the logistic constraints associated with molecular studies. A higher m_e for a given n would, by definition, lower the $n/\hat{m}_{e(p)}$ ratio. Without knowing a colony's m_e a priori, only colonies with low effective paternities or high sample sizes would be above the threshold ratio. Thus, some colonies with effective paternities >15 may have been tested, but there is no current example of an "extreme" effective paternity with an $n/\hat{m}_{e(p)}$ ratio large enough to be 95% certain to estimate m_e to within one subfamily.

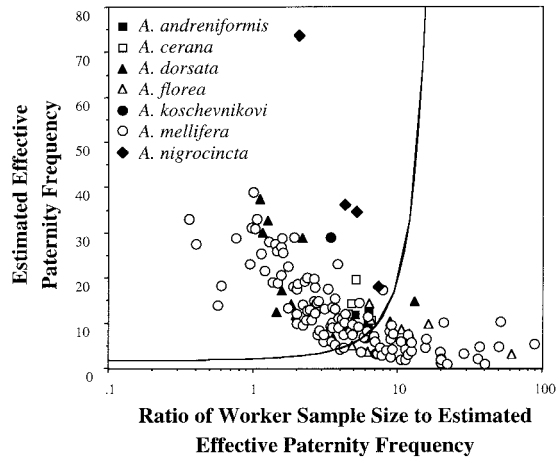


Fig. 3. Review of individual estimates of m_e based on their $n/\hat{m}_{e(p)}$ ratios. The curve denotes the "threshold" value of $n/\hat{m}_{e(p)}$ that minimizes the error of $\hat{m}_{e(p)}$ to within one subfamily. All points above the curve represent colony estimates with sample sizes below the threshold, and all points on or below the curve are those with sample sizes above the threshold. The x-axis is given as a log scale for graphical purposes. All of the estimates of "extreme" effective paternity derive from studies with $n/\hat{m}_{e(p)}$ ratios below the threshold.

Some estimates are closer to their true m_e values than others. Because of this uncertainty, individual 95% CIs were calculated for each colony in the reviewed data set. Confidence intervals could be generated for each colony by running a Monte Carlo simulation for a particular N_o and n. This would assume, however, that $N_o = m_e$, i.e., that all drone fathers were (a) observed in the sampled offspring and (b) uniform in frequency within the brood. Alternatively, confidence intervals of $\hat{m}_{e(p)}$ were generated

for each empirical study by using the standard deviations of $\hat{m}_{e(p)}$ from the simulation results in Fig. 2. Although the confidence limits are not uniform around the means at low sample sizes (see above), the standard deviations can still be used to accurately determine the confidence interval for any given $\hat{m}_{e(p)}$. First, the values of N (2, 5, 10, 20, and 30), n (2–200), and the generated 95% CI were natural log transformed to normally distribute the data. Second, these values were analyzed with a multiple regression to produce the equation:

$$\ln [95\% \text{ CI of } \hat{m}_{e(p)}] = 0.643 - 1.101 \ln(n) + 1.637 \ln(N) \quad [9]$$

This regression equation explains 99.8% of the variance in the 95% CI values and is therefore a very good estimate of the confidence intervals around any $\hat{m}_{e(p)}$ value. The reported n and $\hat{m}_{e(p)}$ values of each study were entered into the regression equation to determine the confidence intervals of each $\hat{m}_{e(p)}$ estimate.

Appendix 1 summarizes the studies of six subspecies and two feral populations of *A. mellifera*. *Appendix* 2 summarizes the studies of six *Apis* species other than *A. mellifera*. The colony number, worker sample size, and $n/\hat{m}_{e(p)}$ ratio is given for each colony in each study. Furthermore, the N_o and m_e estimates (±95% CL) are shown graphically for each colony. The means ± SDs of n, $n/\hat{m}_{e(p)}$, $\hat{m}_{e(p)}$, and N_o are given for each species and subspecies. These values are given for all reports as well as the subset of colonies with $n/\hat{m}_{e(p)}$ values above the "threshold."

Unequal Paternity. Thus far, the simulations have assumed equal representation of the drone fathers in their probability to sire worker offspring. As several studies have pointed out (e.g., Oldroyd et al. 1998), this assumption is the most parsimonious but is most likely to be unrealistic. In fact, 73.5% of the studies with $n/\hat{m}_{e(p)}$ ratios above the threshold have N_o estimates that lie above their 95% CIs of $\hat{m}_{e(p)}$, demonstrating that the majority of studies which minimize sampling error have paternities that are significantly nonuniform. Therefore, it would be helpful to ascertain the effect of unequal subfamily representation on effective paternity estimates.

It would be ideal to perform simulations by randomly sampling workers from a nonuniform distribution of paternities within theoretical colonies. It is impossible, however, to determine a standard distribution that describes the subfamily frequencies in all reported studies (D. Nielsen unpublished data). Instead, the sensitivity of $\hat{m}_{e(p)}$ to non-uniform subfamily proportions was determined by isolating colonies from previous studies with among the highest $n/\hat{m}_{e(p)}$ ratios. These colonies were chosen because they represent the closest approximations to their true, nonuniform subfamily proportions as possible (see above). The same computer simulations described above were used to generate the results, but the program defined the proportions of each subfamily as reported in its respective study instead of assuming uniform male representation in the offspring. The simulation then sampled 2–200 workers from these colonies to calculate the mean $\hat{m}_{e(p)}$ estimates with their 95% CL.

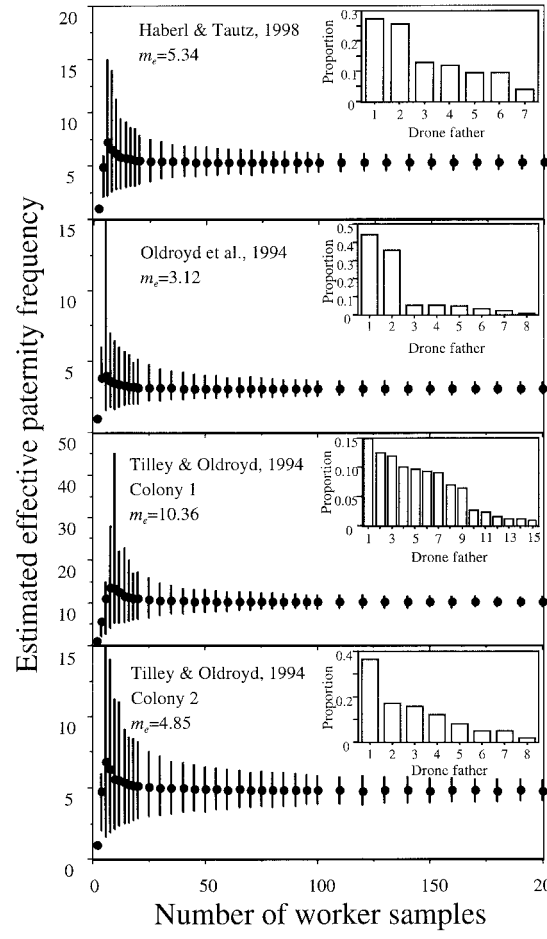


Fig. 4. Effect of non-uniform paternity on estimates of m_e. Results are from simulations that sampled repeatedly workers from four colonies with among the highest $n/\hat{m}_{e(p)}$ ratios. Inset figures illustrate the representation of each drone father within the brood. The errors of $\hat{m}_{e(p)}$ are greater in magnitude compared to those where subfamilies are uniform. ● = $\hat{m}_{e(p)}$ with 95% confidence intervals. Sample sizes that are illustrated: n = 2, 4, 6, . . ., 20, 25, 30, . . ., 100, 110, 120, . . ., 200.

The results demonstrate that the effect of unequal subfamily representation on $\hat{m}_{e(p)}$ is an increased variance in the estimate, which is indicated by the larger 95% CI around the means (Fig. 4). For example, Haberl and Tautz (1998) tested a colony that had an effective paternity frequency of 5.34. This is roughly equivalent to the second panel in Fig. 2 (N = 5). Note that the confidence intervals in Haberl and Tautz (1998) are significantly larger than those when the drone fathers are assumed to be equally represented in the worker offspring.

It appears that sampling error has an even more profound impact on effective paternity estimates of

"natural" (nonsimulated) colonies because of unequal representation among subfamilies, which in some cases can be significant (e.g., Oldroyd et al. 1994). The threshold $n/\hat{m}_{e(p)}$ function is, therefore, the most liberal function possible. In other words, not all studies above the threshold $n/\hat{m}_{e(p)}$ (as it is currently defined) may estimate m_e to within one subfamily because of variation in subfamily representation. Similarly, the 95% CI that were generated for the *Appendices* are the most conservative possible.

Other Factors that May Overestimate Effective Paternity

Phantom Subfamilies. Some workers within a sample may be scored as having marker sets that are different from the true genotypes within a colony. These "phantom" workers may be a result of one or more phenomena. First, a mutation at one or more loci may cause a worker to have a different genotype from her sisters and thus appear like a worker with a different drone father. In particular, mutation rates at microsatellite loci are relatively high, occurring on the order of 10^{-3} to 10^{-6} (Ellegren 2000). We reviewed 19 studies that used microsatellites, which investigated a total of 8,055 workers and scored a minimum of 31,032 alleles. Assuming a mutation rate of 10^{-3}, there may be as many as 31 (0.4%) false marker sets that have been scored because of mutations. Second, the proximity of colonies may enable workers to "drift" between colonies. This phenomenon is exacerbated in domesticated colonies of *Apis mellifera* (Pfeiffer and Crailsheim 1998) and the communal tree nesting species *A. dorsata* (Moritz et al. 1995). Drifters may have common marker sets by descent if adjacent colonies are led by related queens, which may often occur in *A. mellifera* "mating yards" (e.g., Neumann et al. 1999b) but not in *A. dorsata* aggregations (Oldroyd et al. 2000).

The Monte Carlo simulations were used to determine the impact of phantom subfamilies on estimates of m_e. The subfamilies in the program were, again, assumed to be uniform within the offspring, and both paternity number (N = 2, 5, 10, and 20) and worker sample size (n = 10, 20, 30, 40, and 50) were varied independently. One thousand iterations were performed for each N/n combination, and the mean $\hat{m}_{e(p)}$ was estimated with 95% CL as described above. The simulation then produced one or two phantom subfamilies in each data set. Phantom subfamilies were added by choosing randomly one or two workers, respectively, from the sample of n workers, recalculating the new subfamily proportions, and recalculating the $\hat{m}_{e(p)}$ means and confidence limits. Fig. 5 illustrates how $\hat{m}_{e(p)}$ may be even further exaggerated because of sampling error when phantom subfamilies are included in a sample.

This source of error, however, is unlikely to have had a major impact on most estimates of m_e. Phantom subfamilies only have a significant impact on estimates of $\hat{m}_{e(p)}$ at low paternity levels and worker sample sizes (Fig. 5). In order for any worker to be classified as progeny of a given queen, she must share an allele at each locus with her mother. If a mutation occurs in any of the examined maternal alleles, this worker will be excluded as a drifter unless the queen and drone father share an allele at this locus. If the mutation occurs in any paternal alleles, however, the resulting worker will be scored as belonging to an additional "phantom" subfamily. Drifters can be detected easily because of their foreign marker sets (e.g., Moritz et al. 1995) and have been minimized through careful sampling techniques—such as sampling brood (e.g., Oldroyd et al. 1997) or emerging bees from combs (Tarpy and Page 2000)—and increasing the number of loci investigated (Boomsma and Ratnieks 1996). Never-

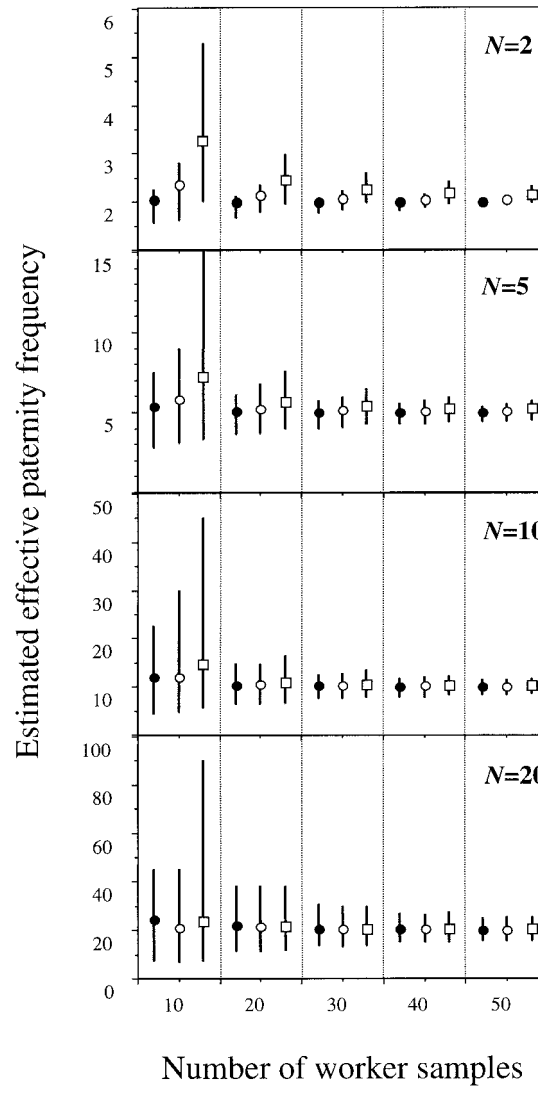


Fig. 5. Effect of "phantom" subfamilies on estimates of m_e. Results are from simulations that sampled 10, 20, 30, 40, and 50 workers from colonies with N equally-represented subfamilies. ● = $\hat{m}_{e(p)}$ with no phantom workers; ○ = $\hat{m}_{e(p)}$ with one phantom worker; □ = $\hat{m}_{e(p)}$ with two phantom workers; all with 95% confidence intervals.

theless, it may be prudent to scrutinize subfamilies that are represented by a single individual, particularly in large data sets, because it is possible that they are not real.

Functional Polygyny. Many social insects may have two or more laying queens within a colony (i.e., are polygynous), but honey bees are exclusively monogynous (reviewed by Crozier and Pamilo 1996). There are brief periods, however, when the offspring of two queens may overlap within honey bee colonies, namely during swarming and supersedure events (Allsopp and Hepburn 1997). More rarely, workers may lay diploid eggs by "self-fertilizing" them by a variety of mechanisms (Laidlaw and Page 1997). Thelytoky is unusually common in the cape bee, *Apis mellifera capensis* (Escholtz), but can also occur in other species (Greeff 1996). If worker offspring are sampled inadvertently from either daughter queens or thelytokous females, they may be mistaken for additional subfamilies of the queen. Both of these events are presumably rare, but nonetheless may serve to overestimate m_e.

The probability of mistakenly including such workers in a queen's brood is a power function of the number and variability of the loci investigated. This probability has been quite low in most studies, because three or more loci have been used to determine paternity. Haberl and Tautz (1999) have developed formulae to estimate the number of laying queens in a colony. The ability to detect a second (or more) egg-laying female is a function of the number of loci, the allelic frequencies at each locus, and the number of worker offspring in the sample. Their study demonstrates that a few workers may have gone undetected, but a large number of workers from an additional queen would have been identified. Even a small number of undetected workers from a second queen, however, could have a large impact on an estimate of m_e if sample size is small.

Mating Ecology. The setting in which queens acquire their mates may be a significant factor in the resultant m_e of colonies. High population densities may likely influence mating frequencies of queens (Haberl and Tautz 1999), such that increased mate availability increases mating frequency. Neumann et al. (1999b) demonstrated that queens in different commercial settings produced different paternity frequencies, which were higher, on average, than those observed in California feral colonies (D. Nielsen unpublished data) where local drone densities are presumed to be much lower than in commercial apiaries. Similarly, because drone densities differ throughout the year (Page 1981), the season that a queen mates may impact her amount of stored sperm (see Camazine et al. 1998) and, presumably, her m_e. Therefore, the source and timing of a particular sample (i.e., commercial versus feral; spring versus fall) may affect estimates of effective paternity frequency. Since the majority of mating estimates for *A. mellifera* have been from commercial settings (at least 92%), it is possible that those estimates may be higher compared with those in a more "natural" context.

Sampling Established Colonies. One final factor that may increase m_e is the state of the colony when it is sampled. Longer-lived colonies have had a greater opportunity to undergo colony-level selection, and it is hypothesized that colonies headed by queens with low insemination frequencies are more subject to colony failure (reviewed by Palmer and Oldroyd 2000). Sampling workers from an established colony therefore may be nonrandom, favoring colonies within the population with higher effective paternities. Furthermore, some studies have sampled swarms (e.g., Rinderer et al. 1998), which may further inflate paternity estimates if higher polyandrous queens have a higher fitness through colony reproduction. This sampling bias may be significant depending on the magnitude of selection acting on paternity frequency. The average $\hat{m}_{e(p)}$ and N_o values in the *Appendices*, therefore, should not be referred to as "the average number of times queens mate" but rather as "the average paternities within tested colonies." To truly understand insemination numbers of queens, paternity tests must be conducted in controlled populations immediately after mating.

In conclusion, our current exploration of honey bee polyandry provides some helpful suggestions for future studies that attempt to quantify the genetic structure of social insect colonies. First, we suggest that sample sizes should be large enough to be above the $n/\hat{m}_{e(p)}$ "threshold" using equation 8. Because m_e is not known a priori, it would be helpful to obtain an initial estimate of m_e, determine the sample size required to meet the threshold, and then run additional samples. Second, we suggest that future estimates of effective paternity use equation 9 to estimate the 95% CI of $\hat{m}_{e(p)}$ so that others may more easily identify the variance in the estimate. Finally, it may help to compare $\hat{m}_{e(s)}$ with $\hat{m}_{e(p)}$ in future studies. If there is a large discrepancy between the two estimates, then it may indicate that sampling error has influenced the result.

The studies reviewed in this paper are the best paternity estimates in any social insect by far and verify that honey bees have among the highest paternity frequencies in the social Hymenoptera. Indeed, 80.5% of the tested colonies have an effective paternity frequency greater than five, and 53.7% have an effective paternity frequency greater than ten. Comparisons among *Apis* species are currently difficult because of the relatively small number of colonies from non-*A. mellifera* species. Nevertheless, *A. nigrocincta* (Smith) colonies have greater paternity frequencies than all other species (all Tukey $q > 2.99$, $P < 0.05$). For the remaining comparisons, only *A. dorsata* has a greater m_e than *A. florea* (F.) ($t = 3.94$, $P < 0.001$), while all other species are not significantly different from one another (all Tukey $q < 2.99$, $P > 0.05$; see also Palmer et al. 2001). The average genetic relatedness values, however, are not significantly different among species (all Tukey $q < 2.99$, $P > 0.05$). All estimates are significantly below $G = 0.50$ (all $t > 1.98$, $P < 0.05$) and approach $G = 0.25$ (Fig. 6). Therefore, polyandry in all *Apis* species have equivalent results, namely a dras-

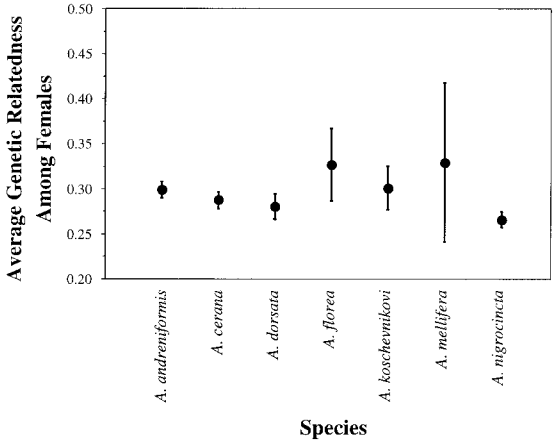


Fig. 6. Estimated average (± SD) intracolonial genetic relatedness within *Apis* colonies.

tic lowering of genetic relatedness among sister nestmates, which is the ultimate factor that concerns kin selection and other social interactions such as worker policing (Moritz et al. 1999; Oldroyd and Osborne 1999), sex allocation Sundström 1994, (Crozier and Pamilo 1996), and division of labor (Calderone et al. 1989; Robinson and Page 1989). Note also that sampling error has profound implications for measuring skews in reproductive groups (e.g., Nonacs 2000).

Many studies of *Apis* do not have large enough sample sizes to be 95% confident that they estimate m_e to within one subfamily. This does not negate the contributions that these studies have made to our understanding of honey bee biology. Neumann et al. (1999b) demonstrated that different mating locations yielded different insemination numbers among queens. Tarpy and Page (2000) showed how insemination numbers on a single mating flight can vary tremendously among queens and that queens may not adjust their mating behaviors in response to their previous insemination number. Neumann and Moritz (2000) illustrate that honey storage may be an increasing function of m_e. The objectives of these studies were not to estimate m_e and G per se but rather to determine the effects of colony genetic structure on different aspects of honey bee biology.

Purposefully missing here is a discussion of factors that may under-estimate m_e, most notably nondetection error (Boomsma and Ratnieks 1996) and differential fertilization over time (Laidlaw and Page 1984). Nondetection error is minimized by increasing the number of loci investigated and has been estimated to have a very low probability of occurring (e.g., Franck et al. 2000). Differential fertilization may be a result of extra- and intraspermathecal events, such as variation in drone sperm contributions, incomplete mixing of sperm, and differential survival of offspring because of disease or genetic incompatibility (Laidlaw and Page 1984). Differences in the short- and long-term paternity frequencies can underestimate m_e by as much as 20% (Laidlaw and Page 1984). Techniques to correct simultaneously for factors that over- and under-estimate m_e have been developed to distinguish single- from double-mated queens (Pedersen and Boomsma 1999), but additional techniques are required for higher paternities.

Enormous strides have been made to describe the extent of polyandry in *Apis.* Nonetheless, caution should be taken with the current effective paternity estimates of honey bee colonies, because it is clear that there are many sources of variation that may affect individual estimates. It is plausible that many, but certainly not all, of the higher estimates may be over-estimates as a result of numerous sampling factors. Therefore, it is best that we take a more conservative view of the extremity of polyandry in *Apis* and temper our astonishment with moderate skepticism.

Examples of extreme polyandry in social insects are attractive systems in which to study kinship theory and social behavior. However, rather than concentrate on the exceptional cases, it may be more helpful in the future to focus on the causes of variation in paternity frequency within and among colonies as well as the impact they have on the evolutionary benefits of polyandry.

Acknowledgments

We thank R. E. Page for his numerous discussions and continuous enthusiasm on this subject. We also thank B. Neff, H. K. Reeve, K. Palmer, B. P. Oldroyd, P. Pamilo, and two anonymous reviewers for providing numerous and valuable comments. The article was written under the USDA postdoctoral fellowship 2001-35302-09905 to DRT.

References Cited

Adams, J., E. D. Rothman, W. E. Kerr, and Z. L. Paulino. 1977. Estimation of the number of sex alleles and queen matings from diploid male frequencies in a population of *Apis mellifera.* Genetics 86: 583–596.

Allsopp, M. H., and H. R. Hepburn. 1997. Swarming, supersedure and the mating system of a natural population of honey bees (*Apis mellifera capensis*). J. Apic. Res. 36: 41–48.

Beye, M., P. Neumann, J. Schmitzova, J. Klaudiny, S. Albert, J. Simuth, M. Felder, and R.F.A. Moritz. 1998. A simple, non-radioactive DNA fingerprinting method for identifying patrilines in honeybee colonies. Apidologia 29: 255–263.

Blanchetot, A. 1991. Genetic relatedness in honeybees as established by DNA fingerprinting. J. Hered. 82: 391–396.

Boomsma, J. J., E. J. Fjerdingstad, and J. Frydenberg. 1999. Multiple paternity, relatedness and genetic diversity in *Acromyrmex* leaf-cutter ants. Proc. R. Soc. Lond. B 366: 219–223.

Boomsma, J. J., and F.L.W. Ratnieks. 1996. Paternity in eusocial Hymenoptera. Phil. Trans. R. Soc. Lond. B 351: 947–975.

Bourke, A.F.G., and N. R. Franks. 1995. Social evolution in ants. Princeton University Press, Princeton, NJ.

Calderone, N. W., G. E. Robinson, and R. E. Page, Jr. 1989. Genetic structure and division of labor in honeybee societies. Experientia 45: 765–767.

Camazine, S., I. Çakmak, K. Cramp, J. Finley, J. Fisher, M. Frazier, and A. Rozo. 1998. How healthy are commer-

cially produced U.S. honey bee queens? Am. Bee J 138: 677–680.

Cornuet, J.-M., and F. Aries. 1980. Number of sex alleles in a sample of honeybee colonies. Apidologie 11: 87–93.

Crozier, R. H. 1970. Coefficients of relationship and the identity of genes by descent in the Hymenoptera. Am. Nat. 104: 216–217.

Crozier, R. H., and P. Pamilo. 1996. Evolution of social insect colonies: sex allocation and kin selection. Oxford University Press, New York.

Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trend. Genet. 16: 551–558.

Estoup, A., M. Solignac, and J. Cornuet. 1994. Precise assessment of the number of patrilines and of genetic relatedness in honeybee colonies. Proc. R. Soc. Lond. B 258: 1–7.

Franck, P., N. Koeniger, G. Lahner, R. M. Crewe, and M. Solignac. 2000. Evolution of extreme polyandry: an estimate of mating frequency in two African honeybee subspecies, *Apis mellifera monticola* and *A. m. scutellata.* Insectes Soc. 47: 364–370.

Gary, N. E. 1963. Observations of mating behaviour in the honeybee. J. Apic. Res. 2: 3–13.

Greeff, J. M. 1996. Thelytokous versus arrhenotokous worker reproduction in the Cape honeybee and other eusocial Hymenoptera. Hereditas 124: 99–103.

Haberl, M., and R.F.A. Moritz. 1994. Estimation of intracolonial worker relationship in a honey bee colony (*Apis mellifera* L.) using DNA fingerprinting. Ins. Soc. 41: 263–272.

Haberl, M., and D. Tautz. 1998. Sperm usage in honey bees. Behav. Ecol. Sociobiol. 423: 247–255.

Haberl, M., and D. Tautz. 1999. Paternity and maternity frequencies in *Apis mellifera sicula.* Insectes Soc. 46: 137–145.

Hamilton, W. D. 1972. Altruism and related phenomena, mainly in social insects. Annu. Rev. Syst. Ecol. 3: 193–232.

Koeniger, G., N. Koeniger, M. Mardan, R.W.K. Punchihewa, and G. W. Otis. 1990. Numbers of spermatozoa in queens and drones indicate multiple mating of queens in *Apis andreniformis* and *Apis dorsata.* Apidologie 21: 281–286.

Koeniger, G., N. Koeniger, and S. Tingek. 1994. Mating flights, number of spermatozoa, sperm transfer and degree of polyandry in *Apis koschevnikovi* (Buttel-Reepen, 1906). Apidologie 25: 224–238.

Kryger, P., and R.F.A. Moritz. 1997. Lack of kin recognition in swarming honeybees (*Apis mellifera*). Behav. Ecol. Sociobiol. 40: 271–276.

Laidlaw, H. H., Jr., and R. E. Page, Jr. 1984. Polyandry in honey bees (*Apis mellifera* L.): sperm utilization and intracolony genetic relationships. Genetics 108: 985–997.

Laidlaw, H. H., Jr., and R. E. Page, Jr. 1997. Queen rearing and bee breeding. Wicwas, Cheshire, UK.

Moritz, R.F.A., P. Kryger, and M. H. Allsopp. 1999. Lack of worker policing in the Cape honeybee (*Apis mellifera capensis*). Behaviour 136: 1079–1092.

Moritz, R.F.A., P. Kryger, G. Koeniger, N. Koeniger, A. Estoup, and S. Tingek. 1995. High degree of polyandry in *Apis dorsata* queens detected by DNA microsatellite variability. Behav. Ecol. Sociobiol. 37: 357–363.

Neumann, P., and R.F.A. Moritz. 2000. Testing genetic variance hypotheses for the evolution of polyandry in the honeybee (*Apis mellifera* L.). Insectes Soc. 47: 271–279.

Neumann, P., R.F.A. Moritz, and J. van Praagh. 1999a. Queen mating frequency in different types of honey bee mating apiaries. J. Apic. Res. 38: 11–18.

Neumann, P., J. P. van Praagh, R.F.A. Moritz, and J. H. Dustmann. 1999b. Testing reliability of a potential island mating apiary using DNA microsatellites. Apidologie 30: 257–276.

Nonacs, P. 2000. Measuring and using skew in the study of social behavior and evolution. Am. Nat. 156: 577–589.

Oldroyd, B. P., M. J. Clifton, K. Parker, S. Wongsiri, T. E. Rinderer, and R. H. Crozier. 1998. Evolution of mating behavior in the genus *Apis* and an estimate of mating frequency in *Apis cerana* (Hymenoptera: Apidae). Ann. Entomol. Soc. Am. 91: 700–709.

Oldroyd, B. P., M. J. Clifton, W. Wongsiri, T. E. Rinderer, H. A. Sylvester, and R. H. Crozier. 1997. Polyandry in the genus *Apis,* particularly *Apis andreniformis.* Behav. Ecol. Sociobiol. 40: 17–26.

Oldroyd, B. P., and K. E. Osborne. 1999. The evolution of worker sterility in honeybees: the genetic basis of failure of worker policing. Proc. R. Soc. Biol. Sci. B 266: 1335–1339.

Oldroyd, B. P., K. E. Osborne, and M. Mardan. 2000. Colony relatedness in aggregations of *Apis dorsata* Fabricius - (Hymenoptera, Apidae). Insectes Soc. 47: 94–95.

Oldroyd, B. P., A. J. Smolenski, J. Cornuet, S. Wongsiri, A. Estoup, T. E. Rinderer, and R. H. Crozier. 1995. Levels of polyandry and intracolonial genetic relationships in *Apis florea.* Behav. Ecol. Sociobiol. 37: 329–335.

Oldroyd, B. P., A. J. Smolenski, J. M. Cornuet, S. Wongsiri, A. Estoup, T. E. Rinderer, and R. H. Crozier. 1996. Levels of polyandry and intracolonial genetic relationships in *Apis dorsata* (Hymenoptera: Apidae). Ann. Entomol. Soc. Am. 89: 276–283.

Oldroyd, B. P., H. A. Sylvester, S. Wongsiri, and T. E. Rinderer. 1994. Task specialization in a wild bee, *Apis florea* (Hymenoptera: Apidae), revealed by RFLP banding. Behav. Ecol. Sociobiol. 34: 25–30.

Page, R. E., Jr. 1981. Protandrous reproduction in honey bees. Environ. Entomol. 10: 359–362.

Page, R. E., Jr. 1986. Sperm utilization in social insects. Annu. Rev. Entomol. 31: 297–320.

Palmer, K., B. Oldroyd, P. Franck, and S. Hadisoesilo. 2001. Very high paternity frequency in *Apis nigrocincta.* Insectes Soc. 48: 327–332.

Palmer, K. A., and B. P. Oldroyd. 2000. Evolution of multiple mating in the genus *Apis.* Apidologie. 31: 235–248.

Palmer, K. A., and B. P. Oldroyd. 2001. Mating frequency in *Apis florea* revisited (Hymenoptera, Apidae). Insectes Soc. 48: 40–43.

Pamilo, P. 1993. Polyandry and allele frequency differences between the sexes in the ant *Formica aquilonia.* Heredity 70: 472–480.

Pamilo, P., and R. H. Crozier. 1982. Measuring genetic relatedness in natural populations: methodology. Theoret. Pop. Biol. 21: 171–193.

Pedersen, J. S., and J. J. Boomsma. 1999. Multiple paternity in social Hymenoptera: estimating the effective mate number in single-double mating populations. Mol. Ecol. 8: 577–587.

Pfeiffer, K. J., and K. Crailsheim. 1998. Drifting of honeybees. Insectes Soc. 45: 151–167.

Queller, D. C., and K. F. Goodnight. 1989. Estimating relatedness using genetic markers. Evolution 43: 258–275.

Rinderer, T. E., J. A. Stelzer, B. P. Oldroyd, and S. Tingek. 1998. Levels of polyandry and intracolonial genetic relationships in *Apis koschevnikovi.* J. Apic. Res. 37: 281–287.

Roberts, W. C. 1944. Multiple mating of queen bees proved by progeny and flight tests. Glean. Bee Cult. 72: 255–260.

Robinson, G. E., and R. E. Page, Jr. 1989. Genetic basis for division of labor in insect societies, pp. 61–80. *In* M. D.

Breed and R. E. Page, Jr. (eds.), The genetics of social evolution. Westview, Boulder, CO.

Sasaki, K., T. Satoh, and Y. Obara. 1995. Sperm utilization by honey bee queens; DNA fingerprinting analysis. Appl. Entomol. Zool. 30: 335–341.

Starr, C. K. 1984. Sperm competition, kinship, and sociality in the Aculeate Hymenoptera, pp. 427–464. *In* R. L. Smith (eds.), Sperm competition and the evolution of animal mating systems. Academic, New York.

Strassmann, J. E. 2001. The rarity of multiple mating by females in the social Hymenoptera. Insectes Soc. 48: 1–13.

Sundström, L. 1994. Sex ratio bias, relatedness asymmetry and queen mating frequency in ants. Nature (Lond.) 367: 266–268.

Taber, S. 1954. The frequency of multiple mating of queen honey bees. J. Econ. Entomol. 47: 995–998.

Taber, S., and J. Wendel. 1958. Concerning the number of times queen bees mate. J. Econ. Entomol. 51: 786–789.

Tarpy, D. R., and R. E. Page, Jr. 2000. No behavioral control over mating frequency in queen honey bees (*Apis mellifera* L.): implications for the evolution of extreme polyandry. Am. Nat. 155: 820–827.

Tilley, C. A., and B. P. Oldroyd. 1997. Unequal subfamily proportions among honey bee queen and worker brood. Anim. Behav. 54: 1483–1490.

West-Eberhard, M. J. 1975. The evolution of social behavior by kin selection. Q. Rev. Biol. 50: 1–33.

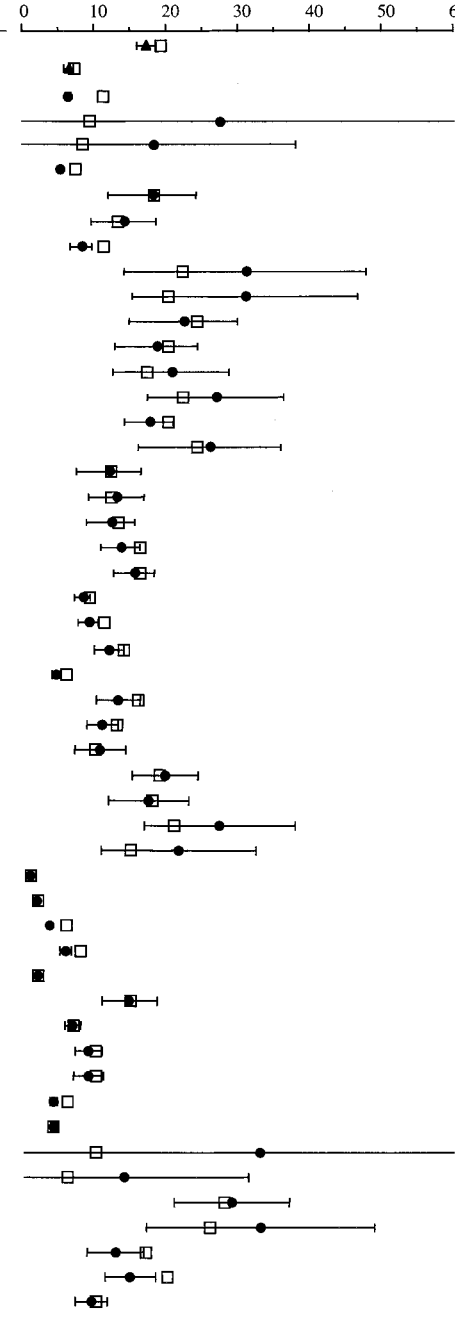
Woyke, J. 1962. Natural and artificial insemination of queen honeybees. Bee World 43: 21–25.

Woyke, J. 1964. Causes of repeated mating flights by queen honeybees. J. Apicult. Res. 3: 17–23.

Received for publication 27 August 2001; accepted 20 February 2002.

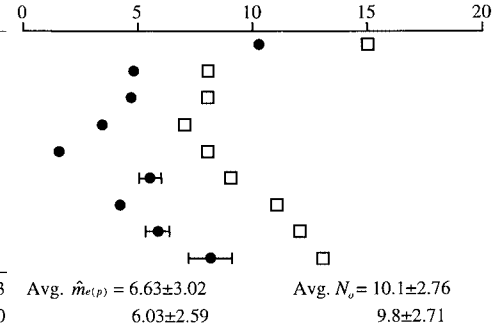
Appendix. Review of individual colony m_e estimates using molecular techniques. N_o and m_e shown graphically for each colony. Error bars for m_e are calculated from equation 9, and are conservative (i.e., smaller) because of the equal-paternity assumption used to generate them (see text). Effective paternity estimates that lack error bars have extremely small confidence limits and do not appear. □ = Reported N_o; ● = $\hat{m}_{e(p)}$, calculated from reported data if available; ▲ = Reported m_e if subfamily proportions were not available to directly calculate $\hat{m}_{e(p)}$, but the values were verified to be equivalent to $\hat{m}_{e(p)}$ (see methods used to calculate m_e). Means ± SD are given for each species and subspecies. Methods for calculating m_e and G in published report: 0 = none or G only; 1 = $\hat{m}_{e(s)}$ and G independently, using equations 1 and 4, respectively; 2 = $\hat{m}_{e(p)}$ and G independently, using equations 2 and 4, respectively; 3 = $\hat{m}_{e(s)} \rightarrow \hat{G}$, using equations 1 and 3, respectively; 4 = $\hat{m}_{e(p)} \rightarrow \hat{G}$, using equations 2 and 3, respectively; 5 = $\hat{G} \rightarrow \hat{m}_{e(s)}$, using equations 4 and 5, respectively; 6 = $\hat{G} \rightarrow \hat{m}_{e(p)}$, using equations 6 and 5, respectively; 7 = $\hat{N}_e = m_e \rightarrow \hat{G}$, using equations 7 and 3, respectively. a = subspecies not given but rather assumed; b = N_o reported as 0; inferred that N_o = 20 from colony 6 in Neumann and Moritz 2000; c = colonies L2-7 and N1-4, 6 are reported in Neumann et al. 1999a; d = reported m_e does not equal $\hat{m}_{e(s)}$ or $\hat{m}_{e(p)}$; e = colonies 1–16 reported in Neumann et al. 1999a; colony 17 not listed in their Table 1; Colony 27 listed twice; f = $\hat{N}_e$ incorrectly calculated; g = only colonies whose queens started to lay eggs after one mating flight (Group ONE) are listed; two colonies with N_o = 0 are not included; proportional data not available in original report; h = brood data only, inferred from their Figure 1; i = claim to calculate $\hat{m}_{e(p)}$ but report $\hat{m}_{e(s)}$; j = m_e reported in table and figures do not correspond; used data from figures because it provided subfamily proportions; k = sum of both randomly-sampled worker and pupae data.

Subspecies	Reference	Colony	n	$n/\hat{m}_{e(p)}$
A. m. carnica (Pollmann)	Estoup et al. 1994[6]	1	137	7.90
		2	59	8.97
	Haberl and Moritz 1994[2, a]	1	102	15.7
	Beye et al. 1998[0, a]	A	11	0.40
		B	11	0.60
	Haberl and Tautz 1998[0]	1	469	87.9
	Neumann et al. 1999a[4, 6]	GW1	34	1.88
		GW2	31	2.20
		GW3	39	4.75
		GW4	30	0.97
		GW5	32	1.03
		GW6	39	1.74
		HG1	38	2.04
		HG2	32	1.55
		HG3	42	1.56
		KW1[b]	58	3.30
		KW2	38	1.46
		KW3	24	2.00
		KW4	32	2.45
		L2	33	2.69
		L3	48	3.53
		L4	57	3.68
		L5	51	6.16
		L6	49	5.38
		L7	52	4.35
		N1	35	7.59
		N2	41	3.10
		N3	43	3.90
		N4	25	2.33
		N6	52	2.63
		RD1	35	2.00
		RD2	39	1.42
		RD3	26	1.20
	Neumann et al. 1999b[6, a, c]	B1	40	40.0
		B2	39	20.0
		B3	72	19.7
		B4	41	6.95
		B5	40	20.0
		B6	39	2.63
		B7	40	5.79
		B8	35	3.88
		B9	32	3.55
		B10	34	8.30
		B11	38	9.19
		L1	12	47.1
		N5[d]	8	0.57
	Neumann and Moritz 2000[4,6,e]	18	56	1.93
		19	35	1.06
		20	30	2.35
		21	42	2.83
		22	32	3.42



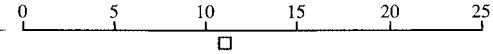
		23[f]	22	0.76	
		24	29	1.14	
		25	28	1.48	
		26[f]	22	0.95	
		27a	26	1.36	
		27b	36	1.29	
		28	32	4.26	
		29	27	2.54	
		30[f]	39	1.00	
		Overall	60	46±58.8	6.1±12.6
		Above threshold	12	84±122.8	20.8±23.2

Avg. $\hat{m}_{e(p)}$ = 15.8±9.23 Avg. N_o = 13.8±6.52
4.4±1.97 5.6±2.87



A. mellifera, feral	Tilley and Oldroyd 1997[0]	1	531	51.3	
		2	243	50.1	
		3	175	36.8	
	D. Nielsen unpublished data	1	72	15.5	
		2	72	35.7	
		3	72	9.38	
		4	72	12.5	
		5	72	8.85	
		6	72	6.28	
		Overall	9	153±145	25.2±18.3
		Above threshold	8	163±162	27.5±18.0

Avg. $\hat{m}_{e(p)}$ = 6.63±3.02 Avg. N_o = 10.1±2.76
6.03±2.59 9.8±2.71



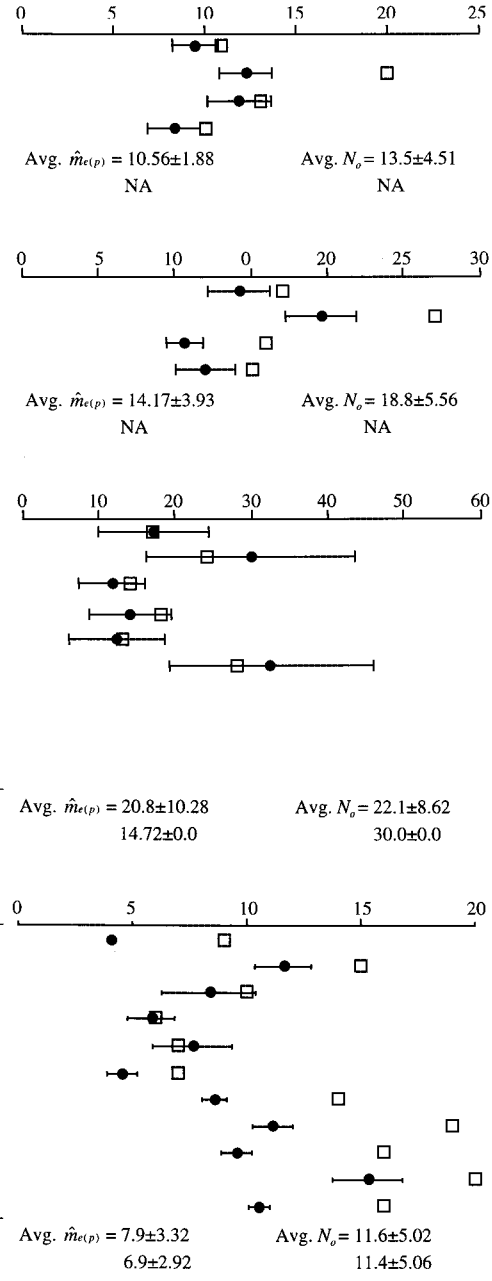
A. m. ligustica (Spinola)	Blanchetot 1991[0, a]	1	141	.	
	Estoup et al. 1994[6]	1	48	2.68	
	Sasaki et al. 1995[0]	1	.	.	
		2	.	.	
	Tarpy and Page 2000[1, g]	5	22	20.0	
		8	24	2.87	
		11	23	7.10	
		12	23	7.82	
		15	23	7.28	
		16	21	2.20	
		19	23	3.09	
		27	21	10.6	
		29	19	3.67	
		33	17	4.00	
		34	20	2.00	
		55	23	8.98	
		61	22	11.3	
		75	23	1.73	
		82	20	5.59	
		84	19	4.22	
		92	18	3.06	
		94	21	2.80	
		96	21	3.50	
		100	22	22.0	
		Overall	22	28±26.0	6.5±5.60
		Above threshold	12	31±34.6	9.9±5.97

Avg. $\hat{m}_{e(p)}$ = 5.78±4.260 Avg. N_o = 6.8±4.26
2.75±1.162 4.4±2.65

A. m. mellifera L.	Estoup et al. 1994[6]	1	211	20.8		
		2	43	4.29		
	Kryger and Moritz 1997[b, h]	A	78	5.40		
		B	87	9.05		
	Overall	4	105±73.3	9.9±7.55	Avg. $\hat{m}_{e(p)}$ = 11.06±2.26	Avg. N_o = 15.2±4.42
	Above threshold	2	149±86.7	14.9±8.3	9.88±0.37	15.0±4.24
A. m. monticola (Smith)	Franck et al. 2000[4]	1	48	3.66		
		2	48	6.85		
		3	48	12.4		
		4	47	11.3		
		5	40	2.77		
		6	95	28.64		
		7	44	2.28		
		8	47	4.83		
	Overall	8	52±17.5	9.1±8.76	Avg. $\hat{m}_{e(p)}$ = 9.37±5.84	Avg. N_o = 11.6±5.01
	Above threshold	4	60±23.7	14.8±9.53	4.59±1.65	7.5±2.38
A. m. scutellata (Lepeletier)	Franck et al. 2000[4]	1	45	1.57		
		2	89	11.9		
		3	89	3.84		
		4	45	4.14		
		5	47	2.35		
		6	41	1.60		
	Overall	6	59±23.1	4.2±3.91	Avg. $\hat{m}_{e(p)}$ = 19.32±8.43	Avg. N_o = 18.8±6.43
	Above threshold	1	89±0.0	11.9±0.0	7.49±0.0	10.0±0.0
A. m. sicula (Grassi)	Haberl and Tautz 1999[5, i]	1	36	4.86		
		2	36	11.8		
		3	62	10.62		
		4	36	3.54		
	Overall	4	42±13.0	7.7±4.10	Avg. $\hat{m}_{e(p)}$ = 6.62±2.969	Avg. N_o = 9.2±2.99
	Above threshold	2	49±18.4	11.2±0.81	4.45±1.966	8.5±4.95
	Overall *A. mellifera* L.	113	54±70.0	8.0±12.2	Avg. $\hat{m}_{e(p)}$ = 12.41±8.754	Avg. N_o = 12.0±6.41
	Above threshold	41	83±108.3	17.6±16.5	4.64±2.449	6.8±3.83

APPENDIX 2

Species	Reference	Colony	n	$n/\hat{m}_{e(p)}$
A. andreniformis (Smith)	Oldroyd et al. 1997[1]	1	60	6.24
		2	78	6.31
		3	60	5.05
		4	41	4.90
	Overall	4	60±15.1	5.62±0.75
	Above threshold	0	NA	NA
A. cerana F.	Oldroyd et al. 1998[3]	1	68	4.78
		2[d]	101	5.14
		3	71	6.63
		4	55	4.56
	Overall	4	74±19.4	5.28±0.93
	Above threshold	0	NA	NA
A. dorsata F.	Moritz et al. 1995[6,7,d]	1930	27	1.57
		1931[j]	35	1.17
		1934[j]	23	1.98
		1935[j]	26	1.84
		1936[j]	18	1.45
		1945	41	1.26
	Oldroyd et al. 1996[1]	1	194	13.2
		2	44	4.84
		3[d]	42	1.12
		4	64	2.22
	Overall	10	51±51.8	3.06±3.72
	Above threshold	1	194±0.0	13.18±0.0
A. florea F.	Oldroyd et al. 1994[0]	B	191[k]	61.2
	Oldroyd et al. 1995[1]	1	71	6.63
		2	24	3.22
		3	24	4.87
		4	24	3.56
		5	23	6.27
	Palmer and Oldroyd 2001[4]	1	94	12.21
		2	92	8.97
		3	93	10.71
		4	93	6.46
		5	159	16.44
	Overall	11	81±56.5	12.7±16.5
	Above threshold	7	96±62.6	17.2±19.8



					0 10 20 30 40 50 60
A. koschevnikovi (Buttel-Reepen)	Rinderer et al. 1998[1]	1	40	4.77	
		2	41	4.30	
		3[d]	40	5.95	
		4[d]	100	3.46	
	Overall	4	55±29.8	4.62±1.04	Avg. $\hat{m}_{e(p)}$=13.40±10.43 Avg. N_o=16.2±10.53
	Above threshold	1	40±0.0	5.95±0.0	6.72±0.0 10.0±0.0

					0 10 20 30 40 50 60 70 80 90
A. nigrocincta (Smith)	Palmer et al. 2001[4]	1	163	4.52	
		2	181	5.26	
		3	154	2.09	
		4	139	7.80	
	Overall	4	159±17.6	4.9±2.35	Avg. $\hat{m}_{e(p)}$ = 40.5±23.58 Avg. N_o = 54.0±11.52
	Above threshold	0	NA	NA	NA NA