

# A Bayesian Model for Assessing the Frequency of Multiple Mating in Nature

B. D. Neff, T. E. Pitcher, and J. Repka

Many breeding systems have multiple mating, in which males or females mate with multiple partners. With the advent of molecular markers, it is now possible to detect multiple mating in nature. However, no model yet exists to effectively assess the frequency of multiple mating ( $f_{mm}$ )—the proportion of broods with at least two males (or females) genetically contributing—from limited genetic data. We present a single-sex model based on Bayes' rule that incorporates the numbers of loci, alleles, offspring, and genetic parents. Two genetic criteria for calculating  $f_{mm}$  are considered: the proportion of broods with three or more paternal (or maternal) alleles at any one locus and the total number of haplotypes observed in each brood. The former criterion provides the most precise estimates of  $f_{mm}$ . The model enables the calculation of confidence intervals and allows mutations (or typing errors) to be incorporated into the calculation. Failure to account for mutations can result in overestimates of  $f_{mm}$ . The model can also utilize other biological data, such as behavioral observations during mating, thereby increasing the accuracy of the calculation as compared to previous models. For example, when two sires contribute equally to multiply mated broods, only three loci with five equally common alleles are required to provide estimates of  $f_{mm}$  with high precision. We demonstrate the model with an example addressing the frequency of multiple paternity in small versus large clutches of the endangered Kemp's Ridley sea turtle (*Lepidochelys kempi*) and show that females that lay large clutches are more likely to have multiply mated.

The discovery that multiple mating is prevalent in the animal kingdom has revolutionized the study of mating systems (Jennions and Petrie 2000; Krebs and Davies 1997; Reynolds 1996). Multiple mating occurs when individuals of one sex mate with more than one individual of the opposite sex (Neff et al. 2000a; Reynolds 1996). Here, we define the *frequency* of multiple mating as the proportion of broods in a population that contain genes from at least two males or two females. Multiple mating is typically detected by the presence of three paternal (or maternal) alleles in the young (Kelly et al. 1999; Milkman and Zeitler 1974; Ochando et al. 1996; Zane et al. 1999). However, this approach is conservative, because not all multiply mated broods will be detected—for example, when two males have identical genotypes or are homozygous for alternative alleles, or when insufficient offspring are sampled to detect all of the paternal alleles. When less than three paternal alleles are

detected, several models have been developed to determine if the brood is actually multiply mated. When two alleles are detected, multiple mating can be inferred using either of two approaches: (1) the frequency of the two alleles differs significantly from the 1:1 ratio expected under Mendelian inheritance and assuming a single heterozygous sire (Barry et al. 1992; Birdsall and Nash 1973; Schwartz et al. 1989); or (2) the probability is sufficiently high (e.g.,  $P > .95$ ) that two (or more) sires would have only the two alleles detected, based on frequencies of the alleles in the population (Milkman and Zeitler 1974; Neff and Pitcher 2002; Pedersen and Boomsma 1999; Williams and Evarts 1989). When just one paternal allele is detected, only the latter approach is appropriate. However, these approaches assume a priori that multiply mated broods are just as likely as singly mated broods, an assumption that can lead to inaccurate assessment.

From the Department of Biology, Biological & Geological Sciences Building, University of Western Ontario, London N6A 5B7, Ontario, Canada (Neff); the Department of Zoology, University of Toronto, 25 Harbord Street, Toronto M5S 3G5, Ontario, Canada (Pitcher); and the Department of Mathematics, University of Toronto, 100 St. George Street, Toronto M5S 3G5, Ontario, Canada (Repka). We are grateful to N. Mrosovsky and H. Rodd for helpful discussion and to four anonymous reviewers for their comments. We thank Rene Marquez for providing the turtle clutch size data. This work was supported by the Natural Science and Engineering Council of Canada. Address correspondence to Bryan D. Neff at the address above, or e-mail: bneff@uwo.ca.

© 2002 The American Genetic Association 93:406–414

**Table 1. Summary of the variables used in the models**

Variable	Description
$A_{bl}$	Number of different paternal alleles observed in brood $b$ at locus $l$
$B$	Number of broods analyzed (indexed by $b$ )
$C_b$	Number of offspring analyzed from brood $b$
$F$	Mean number of sires contributing to a brood (for multiple paternity $F > 1$ ; otherwise $F = 1$ )
$f_{mm}$	Frequency of multiple mating
$\mathbf{H}$	Vector containing each value of $H_b$ for all broods
$H_b$	Number of different haplotypes observed among the $C_b$ offspring from brood $b$
$L$	Number of loci used to genotype genetic parent and brood
$M$	Proportion of pseudo-broods that have same number of haplotypes as $H_b$
$Pat_i$	Breeding success of sire $i$
$P_{la}$	Population allele frequency of allele $a$ at locus $l$
$P_m$	Observed proportion of broods with three or more paternal (or maternal) alleles

We have previously developed a model that calculates the statistical power of genetic analyses to detect multiple mating, based on detecting three or more paternal alleles within an offspring sample (e.g., a brood) (Neff and Pitcher 2002). We now place this model in a Bayesian framework and develop a model to calculate the frequency of multiple mating from a set of broods. This framework explicitly defines the prior probability of multiple mating and thereby avoids assumptions made by previous models. The model can incorporate genetic and other biological data as well as mutations (or typing errors). The model can also be used to calculate confidence intervals associated with each estimate. We focus on mating systems with single-sex multiple mating. Single-sex multiple mating occurs when there is multiple mating by only one sex. Here, we define multiple mating at the level of the brood (Neff et al. 2000a). Therefore, single-sex multiple mating gives rise to broods that contain the genes of either a single female that has mated with multiple males (genetic polyandry) or a single male that has mated with multiple females (genetic polygyny). All of the young within a brood are either full or half sibs. Our model requires single-locus codominant genetic data, such as microsatellites or allozymes, from the genetic parent and a sample of the brood. A genetic sample from the breeding population is also required to estimate allele frequencies.

We consider two genetic criteria in our model: (1) the proportion of broods analyzed in which there are three or more paternal (or maternal) alleles detected at any one locus, and (2) the number of haplotypes detected in each brood. The second criterion has been shown to be more sensitive at detecting the actual number of mates contributing to multiply mated broods (DeWoody

et al. 2000a,b). For the second criterion we assume either full recombination (i.e., each locus segregates independently) or no recombination. We show that seven factors affect the accuracy and precision of the estimates of the frequency of multiple mating: (1) the genetic criterion used in the model, (2) the number of loci, (3) the number of alleles and their frequencies, (4) the number of offspring analyzed from the brood, (5) the number of broods analyzed, (6) the number of genetic parents and the reproductive skew among the genetic parents, and (7) the prior probability of multiple mating.

We demonstrate our model with a biological example addressing multiple paternity in the endangered Kemp's Ridley sea turtle (*Lepidochelys kempi*). Female Kemp's Ridley sea turtles may mate with one or more males prior to nesting and can lay multiple clutches within a single nesting season, but they do not appear to mate between clutches (Mendonça and Pritchard 1986). Turtles are capable of storing sperm for long periods of time, and therefore multiple paternity likely arises from sperm stores (Kichler et al. 1999). To investigate the frequency of multiple mating in Kemp's Ridley sea turtles, Kichler et al. (1999) used up to three microsatellite loci to genotype clutches from 26 females. They developed a maximum likelihood model to test whether each clutch was fertilized by a single or two males and a second model to test for a skew in paternity, given a multiply mated clutch by two males. They showed that the most likely solution is that all clutches were multiply sired (by the assumed two males), with skewed paternity of 0.753 and 0.247. However, they could not reject the possibility that as few as 70% of clutches were multiply sired. Kichler et al. also found that multiple paternity was more frequently detected in larger clutches

and qualitatively attributed this to the increased likelihood of detecting three or more paternal alleles in the larger samples. They did not consider the alternative that females that lay larger clutches mate with multiple males more frequently than females that lay smaller clutches (Trexler et al. 1997). We use our model to calculate the frequency of multiple paternity and compare this estimate to the estimate provided by Kichler and colleagues. Next, we investigate whether the frequency of multiple paternity is lower in small (<160 eggs, the median number of eggs) versus large ( $\geq 160$  eggs) clutches, and we use an exact statistic based on the 95% confidence intervals associated with the estimates to determine if large clutches (and hence larger females) are more likely to multiply mate.

## Methods

### The Model

All of the variables used in the model are defined in Table 1. Our model is based on Bayes' rule, which provides the appropriate framework to calculate conditional probabilities and, hence, the frequency of multiple mating from limited biological data (Lewis 2001; Neff et al. 2001). Suppose that genetic data has been obtained from a sample of  $B$  broods and that the single genetic mother is known for each brood. By subtracting the mother's genetic contribution, the minimum number of different paternal alleles and haplotypes can be constructed from the offspring. The maximum number of paternal alleles that any one sire can contribute to a brood is two per locus, and the maximum number of haplotypes across  $L$  loci is  $2^L$  if there is full recombination (i.e., loci segregate independently) or 2 if there is no recombination (Table 2).

First, consider the number of different paternal alleles, and let  $A_{bl}$  represent the number of different paternal alleles detected in brood  $b$  at locus  $l$ . A multiply mated brood is detected when  $A_{bl} > 2$  for a given  $b$  and any  $l$ . Next, define  $P_m$  as the proportion of  $B$  broods in which a multiple mating is detected. The actual frequency of multiple mating within a population may be higher than  $P_m$ , because a brood may be multiply mated even when two or fewer alleles are detected at each locus. This can occur when, for example, not all sires are heterozygous at each locus, some sires have identical alleles, or only a single

**Table 2. Minimum number of paternal haplotypes, identified from the genotypes of the known genetic parent and the offspring within its brood**

Known parent		Offspring		Haplotypes <sup>a</sup>		Min. no. of paternal haplotypes <sup>a</sup>	
Locus 1	Locus 2	Locus 1	Locus 2	Independent	Dependent	Independent	Dependent
A <sub>1</sub> A <sub>1</sub>	B <sub>1</sub> B <sub>1</sub>	A <sub>1</sub> A <sub>1</sub>	B <sub>1</sub> B <sub>1</sub>	A <sub>1</sub> B <sub>1</sub>	A <sub>1</sub> B <sub>1</sub>	4	4
		A <sub>1</sub> A <sub>1</sub>	B <sub>1</sub> B <sub>2</sub>	A <sub>1</sub> B <sub>2</sub>	A <sub>1</sub> B <sub>2</sub>		
		A <sub>1</sub> A <sub>2</sub>	B <sub>1</sub> B <sub>1</sub>	A <sub>2</sub> B <sub>1</sub>	A <sub>2</sub> B <sub>1</sub>		
		A <sub>1</sub> A <sub>2</sub>	B <sub>1</sub> B <sub>2</sub>	A <sub>2</sub> B <sub>2</sub>	A <sub>2</sub> B <sub>2</sub>		
A <sub>1</sub> A <sub>2</sub>	B <sub>1</sub> B <sub>2</sub>	A <sub>1</sub> A <sub>1</sub>	B <sub>1</sub> B <sub>1</sub>	A <sub>1</sub> B <sub>1</sub>	A <sub>1</sub> B <sub>1</sub>	2	A <sub>1</sub> B <sub>1</sub> & A <sub>2</sub> B <sub>2</sub>
		A <sub>1</sub> A <sub>2</sub>	B <sub>1</sub> B <sub>2</sub>	A <sub>1</sub> B <sub>1</sub> or A <sub>1</sub> B <sub>2</sub> or A <sub>2</sub> B <sub>1</sub> or A <sub>2</sub> B <sub>2</sub>	A <sub>1</sub> B <sub>1</sub> or A <sub>2</sub> B <sub>2</sub>		
		A <sub>1</sub> A <sub>2</sub>	B <sub>2</sub> B <sub>2</sub>	A <sub>1</sub> B <sub>2</sub> or A <sub>2</sub> B <sub>2</sub>	A <sub>1</sub> B <sub>2</sub>		
		A <sub>2</sub> A <sub>2</sub>	B <sub>1</sub> B <sub>2</sub>	A <sub>2</sub> B <sub>1</sub> or A <sub>2</sub> B <sub>2</sub>	A <sub>2</sub> B <sub>1</sub>		

Examples include two known parents and four offspring from each of their broods genotyped with two loci. In some cases, paternal haplotypes may be ambiguous.

<sup>a</sup> Independent assumes full recombination between loci and dependent assumes no recombination.

sire's offspring are sampled from a nest (due to incomplete sampling). The actual frequency of multiple mating also may be lower than  $P_m$ . This can occur from sampling error introduced when not all broods are analyzed from a population or when a mutation arises in an offspring from a brood of full sibs. Using Bayes' rule, the probability of a certain frequency of multiple mating ( $f_{mm}$ ) given the observed  $P_m$  can be calculated from:

$$\Pr(f_{mm} | P_m) = \frac{\Pr(P_m | f_{mm}) \times \Pr(f_{mm})}{\Pr(P_m)}, \quad (1)$$

where  $f_{mm}$  is an index to the possible frequencies of multiple mating ( $0 \leq f_{mm} \leq 1$ );  $\Pr(f_{mm} | P_m)$  is the conditional probability of a frequency of multiple mating given the proportion of broods with three or more paternal alleles;  $\Pr(P_m | f_{mm})$  is the reverse: the probability of observing that proportion given the frequency of multiple mating  $f_{mm}$ ;  $\Pr(f_{mm})$  is the prior probability of the frequency of multiple mating (independent of the genetic data); and  $\Pr(P_m)$  is the probability of observing the proportion of broods with three or more paternal alleles and represents a normalization constant such that  $\int \Pr(f_{mm} | P_m) df_{mm} = 1$ .

For the second genetic criterion, the number of paternal haplotypes, we do not consider the proportion of broods with greater than  $2^L$  haplotypes, because with even a moderate number of loci multiply mated broods rarely exceed this value (assuming full recombination). It is therefore of limited value. We instead define a vector  $\mathbf{H}$ , containing  $B$  elements, representing the number of haplotypes observed at each brood. We thus consider the actual distribution of haplotypes across all broods. An expression similar to Equation 1 can be developed to calculate the probability of a certain

frequency of multiple mating, given this distribution:

$$\Pr(f_{mm} | \mathbf{H}) = \frac{\Pr(\mathbf{H} | f_{mm}) \times \Pr(f_{mm})}{\Pr(\mathbf{H})}. \quad (2)$$

Equation 2 differs from Equation 1 only in the genetic criterion used. While Equation 1 examines the proportion of broods that are multiply mated, Equation 2 examines the specific distribution of haplotypes at each brood. Given these equations, an unbiased estimate of the actual frequency of multiple mating can be calculated from (Neff et al. 2001):

$$\overline{f_{mm}} = k_1 \cdot \int_0^1 (\Pr(f_{mm} | P_m) \times f_{mm}) df_{mm} \quad (3)$$

or

$$\overline{f_{mm}} = k_2 \cdot \int_0^1 (\Pr(f_{mm} | \mathbf{H}) \times f_{mm}) df_{mm}, \quad (4)$$

where the line above the  $f_{mm}$  denotes the expected value, and  $k_i$  is the normalization constant defined such that  $\int \Pr(f_{mm} | P_m) df_{mm} = 1$  or  $\int \Pr(f_{mm} | \mathbf{H}) df_{mm} = 1$ . Confidence intervals can be established for the estimate from Equation 3 (or Equation 4) by determining the values of  $f_{mm}$  (denoted below with an asterisk) that cut off upper and lower "tails" of areas  $1 - \alpha/2$  and  $\alpha/2$ , respectively, from:

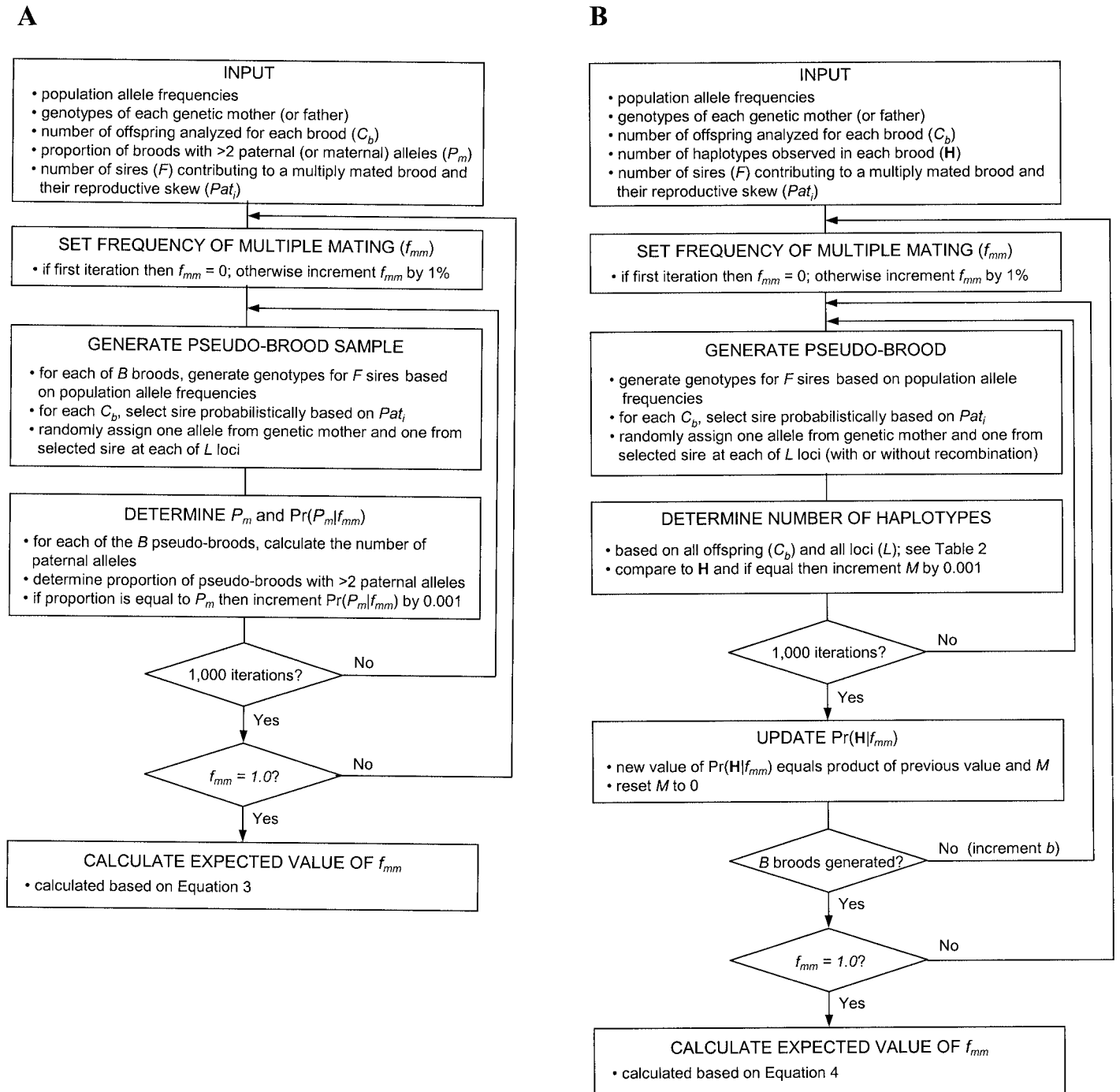
$$k_1 \cdot \int_0^{f_{mm}^*} \Pr(f_{mm} | P_m) df_{mm} = \frac{\alpha}{2} \text{ or } 1 - \frac{\alpha}{2}. \quad (5)$$

As an example, for the 95% confidence interval,  $\alpha$  equals 0.05, and Equation 5 is solved for values of  $f_{mm}$  that cut off the lower and upper 2.5% of the normalized  $\Pr(f_{mm} | P_m)$  distribution. An analogous expression can be developed for  $\mathbf{H}$ .

To calculate the expected frequency of multiple mating using either Equation 3

or 4, the reverse probability [i.e.,  $\Pr(P_m | f_{mm})$  or  $\Pr(\mathbf{H} | f_{mm})$ ] must first be calculated for each possible value of  $f_{mm}$ . The reverse probability will depend on all of: (1) the number of loci, (2) the number of alleles and their frequencies, (3) the number of offspring analyzed from each brood, (4) the number of genetic parents contributing to each brood, and (5) the reproductive skew among the genetic parents. An exact formulation of  $\Pr(P_m | f_{mm})$  or  $\Pr(\mathbf{H} | f_{mm})$  is therefore complex and difficult to evaluate (Neff and Repka, unpublished data). However, Monte Carlo algorithms (Lewis 2001; Manly 1997) can be developed and provide an effective means to estimate its value (Figure 1). Our algorithms are similar to the one developed by Kichler et al. (1999), but allow the number of sires and reproductive skew to be defined a priori. Thus, any number of sires and any level of skew can be considered. Our algorithms similarly enable mutation (and typing error) rates to be incorporated into the analysis.

The prior probability  $\Pr(f_{mm})$  must also be calculated. This can be done from other biological data, such as behavioral observations during mating or previous independent genetic analyses. In Neff et al. (2001) we provide formulas to calculate prior probabilities from biological data. Generally, if nothing is known about the actual distribution of  $\Pr(f_{mm})$ , then assuming a uniform distribution is the least biased (see analysis and discussion in Neff et al. 2001). In this case, the prior probability is independent of  $f_{mm}$ , and it becomes part of the normalization constant  $k_i$  in Equation 3 or 4. Finally, it is unnecessary to explicitly calculate the value of  $\Pr(P_m)$  or  $\Pr(\mathbf{H})$ , because they are independent of  $f_{mm}$  and become part of the normalization constant  $k_i$ .



**Figure 1.** Schematic of the Monte Carlo simulation used to calculate the reverse probability based on (A) the proportion of broods with more than two paternal (or maternal) alleles, or (B) the number of haplotypes observed in each brood.

### The Genetic Criteria

To examine the effects of each genetic criterion on the accuracy and precision of the estimates of  $f_{mm}$ , we conducted the following simulations (Table 3). First, we assumed that either 5, 15, or 50 offspring from each of 5 or 15 broods were sampled from a population. The broods were assumed to be very large, and therefore the sampled offspring repre-

sented only a small proportion of the brood. The frequency of multiple mating was, on average,  $\mu = 0.5$  and followed the (normal) prior probability distribution:

$$\Pr(f_{mm}) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\{-\frac{(f_{mm}-\mu)^2}{2\sigma^2}\}}, \quad (6)$$

where  $\sigma$  (the standard deviation) was set to 0.25. In the event of a multiple mating, it was assumed that either two of five

sires contributed with equal probability, or five sires contributed with relative paternities of 0.516, 0.258, 0.129, 0.065, and 0.032 (i.e., each male had 50% the paternity of the preceding male). We also considered one simulation with very high skew: two sires with 95% and 5% paternity. The skew distribution was assumed to be binomial (two sires) or multinomial (five sires) (i.e., sampling error was

**Table 3. Parameters used in the simulations (Sim) to test the effectiveness of the two genetic criteria and the bias introduced by assuming the incorrect prior probability distribution for the frequency of multiple mating**

Sim #	Sires	Skew <sup>a</sup>	Broods	Offspring	Loci	Alleles	$f_{mm}^b$		Mutation rate <sup>c</sup>	Prior <sup>d</sup>
							$\mu$	$\sigma$		
Genetic criteria analysis										
1	2	Even	15	5	3	5	0.50	0.25	—	Correct
2	2	Even	15	15	3	5	0.50	0.25	—	Correct
3	2	Even	5	15	3	5	0.50	0.25	—	Correct
4	2	Even	15	15	3	20	0.50	0.25	—	Correct
5	5	Even	15	15	3	20	0.50	0.25	—	Correct
6	2	Even	15	15	1	5	0.50	0.25	—	Correct
7	2	Even	15	15	8	5	0.50	0.25	—	Correct
8	5	Skew	15	15	3	20	0.50	0.25	—	Correct
9	2	High	15	15	3	5	0.50	0.25	—	Correct
10	2	High	15	50	3	5	0.50	0.25	—	Correct
11	2	Even	15	15	1	10 (5) <sup>e</sup>	0.50	0.25	—	Correct
Mutation analysis										
12	2	Even	15	15	3	5	0.50	0.25	0.1 (no)	Correct
13	2	Even	15	15	3	5	0.50	0.25	0.1 (yes)	Correct
14	2	Even	15	15	3	5	0.50	0.25	0.01 (no)	Correct
15	2	Even	15	15	8	20	0.50	0.25	0.01 (no)	Correct
16	2	Even	15	15	3	5	0.50	0.25	0.001 (no)	Correct
17	2	Even	15	15	8	20	0.50	0.25	0.001 (no)	Correct
Prior probability analysis										
18	2	Even	15	10	2	5	0.10	0.10	—	Correct
19	2	Even	15	10	2	5	0.10	0.10	—	Uniform
20	2	Even	15	10	2	5	0.90	0.10	—	Correct
21	2	Even	15	10	2	5	0.90	0.10	—	Uniform

<sup>a</sup> Reproductive skew: even denotes equal breeding success for each male, skew denotes each subsequent male has 50% the breeding success of the former male, and high denotes 95% and 5% paternity for each male (see text).

<sup>b</sup> The prior probability of  $f_{mm}$  was assumed to follow a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ .

<sup>c</sup> The yes/no denotes whether or not the mutation rate was incorporated into the calculation of reverse probability.

<sup>d</sup> The prior probability distribution assumed in the analysis (either the correct normal distribution or the incorrect uniform distribution).

<sup>e</sup> Ten census alleles: 1 with frequency 0.4 and 9 with frequency 1/15. Calculated from Neff et al. (2000b), there are five effective alleles.

included), and the correct distribution was used in the simulation and analysis.

Genotypes were generated for each offspring at  $L$  loci, each having either five or 20 equally common alleles. The number of haplotypes under assumptions of either full recombination or no recombination was determined for each brood as well as the proportion of broods with three or more paternal alleles. The estimated frequency of multiple mating was then calculated according to Equations 3 and 4, using the Monte Carlo algorithms (Figure 1) and assuming the correct prior probability distribution (i.e., Equation 6). The simulation was repeated 100 times, and the mean difference (bias) between the estimated and actual frequency of multiple mating was calculated. The associated 95% confidence interval (based on the 100 samples) was also calculated.

### Mutations

To determine the effects of mutations on the estimates of  $f_{mm}$ , we conducted six simulations, following the structure out-

lined above (Table 3). Three mutation rates were considered: 0.1, 0.01, and 0.001. The highest mutation rate was selected to elucidate the potential effects of mutations, and the latter two were selected as potential upper limits for microsatellite loci (Jarne and Lagoda 1996). When a mutation occurred, the allele in the offspring was randomly changed to one of the other alleles at the locus with equal probability. Only mutations in the paternal germ line were considered, to ensure that the offspring matched the known mother at each locus. For one simulation (Sim 13), we included the mutation rate (and process) in our model by incorporating it into the Monte Carlo simulation that calculates the reverse probability  $\Pr(P_m|f_{mm})$ . By including the mutation rate in the latter calculation, we tested whether the model could account for the effects of mutations. The estimated frequency of multiple mating was then calculated, using Equations 3 and 4, as above. The simulation was repeated 100 times, and the mean difference (bias) and the 95%

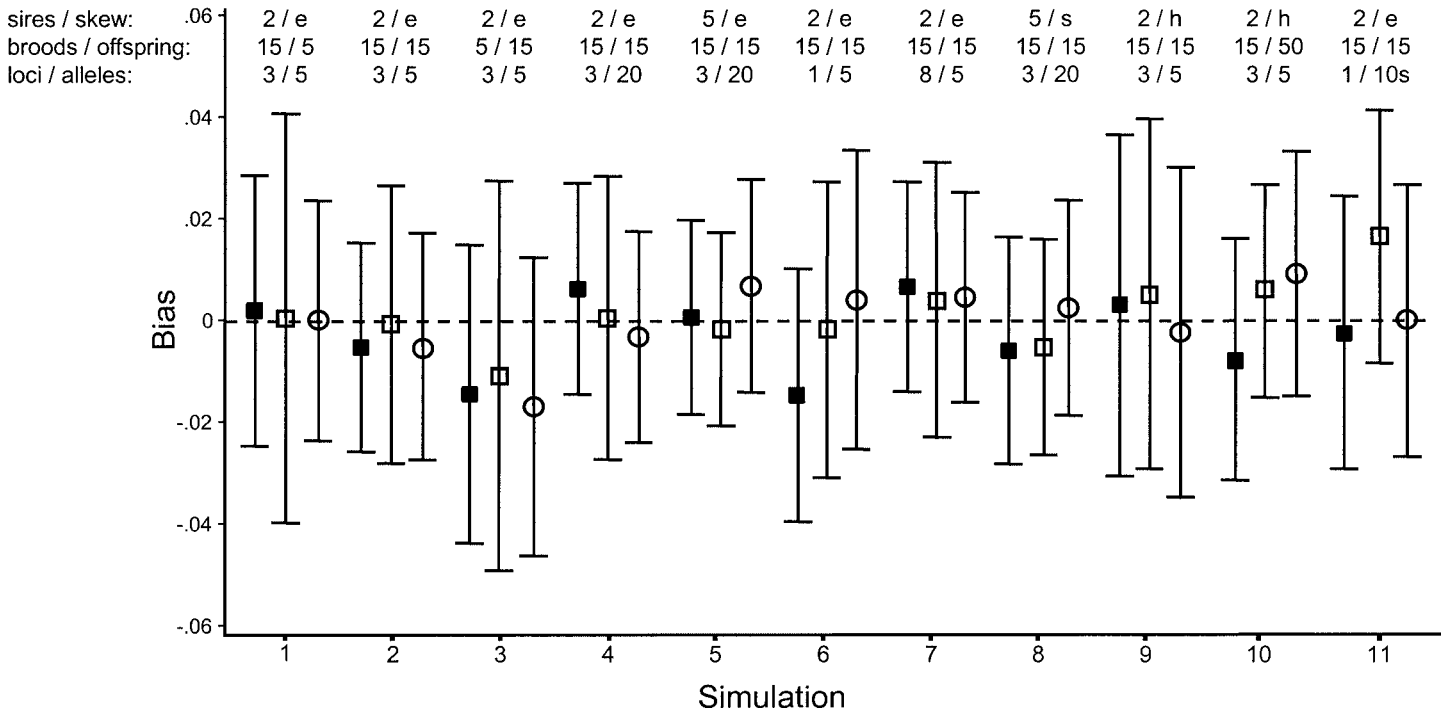
confidence interval between the estimated and actual frequency of multiple mating were calculated.

### The Prior Probability $\Pr(f_{mm})$

To demonstrate the potential influence of the prior probability on the estimate of  $f_{mm}$ , we conducted two simulations, following the structure outlined above (Table 3). In this case, the prior probability of the frequency of multiple mating was assumed to follow the (normal) distribution defined by Equation 6, where  $\mu$  was equal to either 0.10 or 0.90 and  $\sigma = 0.25$ . The frequency of multiple mating was calculated according to Equations 2 and 3, assuming the  $\Pr(f_{mm})$  followed the correct distribution (i.e., Equation 6) or the uniform distribution defined by  $\Pr(f_{mm}) = 1$ . The simulation was repeated 100 times, and the mean difference (bias) and the 95% confidence interval between the estimated and actual frequency of multiple mating were calculated.

### Kemp's Ridley Sea Turtles

Next, we applied our model to address the frequency of multiple paternity between small and large clutches from a population of endangered Kemp's Ridley sea turtles. Using the genetic data presented in Kichler et al. (1999), we quantified the number of unique paternal alleles detected in each clutch. The frequency of multiple mating in the population was first calculated using all clutches. We then classified the clutches as either small (<160 eggs, the median number of eggs) or large ( $\geq 160$  eggs). Because the number of sires contributing to a multiply mated clutch was unknown, we considered four scenarios: (1) two sires with equal breeding success (mean paternity of 50% for each male), (2) two sires with skewed success (66.7% and 33.3%), (3) three sires with equal success (33.3% for each male), and (4) three sires with skewed success (57%, 28.5%, and 14.5%). Furthermore, because the prior probability distribution (i.e.,  $\Pr(f_{mm})$ ) was unknown, we assumed it followed the uniform distribution. This may be the least biased approach in the absence of additional information (Neff et al. 2001). For each scenario, the expected frequency of multiple mating and 95% confidence interval were calculated using the first genetic criterion and assuming a mutation rate of 0.001 (Jarne and Lagoda 1996; Pearse et al. 2002). From the confidence distributions (estimated using higher-order polynomials) the exact probability



**Figure 2.** Simulation results for the genetic criteria analysis. The mean bias and 95% confidence are shown for simulations 1–11. The results from the first genetic criterion (proportion of broods with three or more paternal alleles) are on the left (closed square) of each simulation trio, those for the second criterion (number of haplotypes) with full recombination are in the middle (open square), and those for the second criterion without recombination are on the right (open circle). The numbers above the bars correspond to the parameters listed in Table 3.

that the frequency of multiple mating was higher in large clutches ( $Lg$ ) as compared to small clutches ( $Sm$ ) was calculated from:

$$\begin{aligned} & \Pr(f_{mm}^{Lg} > f_{mm}^{Sm}) \\ &= \int_{f_{mm}^{Lg}=0}^1 \Pr(f_{mm}^{Lg} | P_m^{Lg}) \\ & \times \left( \int_{f_{mm}^{Sm}=0}^{f_{mm}^{Sm} < f_{mm}^{Lg}} \Pr(f_{mm}^{Sm} | P_m^{Sm}) df_{mm}^{Sm} \right) df_{mm}^{Lg}. \end{aligned} \quad (7)$$

## Results

### The Genetic Criteria

For all simulations where the correct prior probability distribution was used (Table 3), both the proportion of broods containing three or more paternal alleles ( $P_m$ ) and the number of paternal haplotypes ( $H$ ) provided unbiased estimates of the actual frequency of multiple mating (i.e., the mean bias did not differ significantly from zero;  $P > 0.50$  for all). Generally,  $P_m$  performed as well or better than  $H$  (with or without recombination), as measured by the precision of the estimates (Figure 2).  $H$  performed best when there was no recombination between loci and provide the most precise estimates of  $f_{mm}$  when there were only

a few offspring analyzed. When a large number of sires contributed to a multiply mated brood and a large number of offspring were analyzed from each brood, the criteria performed about equally well (e.g., Sim 5, 8, and 10).

Increasing the number of offspring sampled from each brood or increasing the number of broods analyzed from a population, each increased the precision of the estimates of the frequency of multiple mating (compare Sim 1 to 2, Sim 3 to 2, and Sim 9 to 10). Thus, better estimates can be obtained by increasing the number of offspring analyzed and the number of broods sampled. Interestingly, loci with a greater number of alleles did not provide much better estimates than did loci with fewer alleles (compare Sim 2 to 4), and loci with similar numbers of effective alleles (but different numbers of census alleles) performed similarly (compare Sim 6 to 11). Thus, only loci with a moderate number of effective alleles (e.g., five) are needed. On the other hand, increasing the number of loci used did increase precision (compare Sim 6 to 2). However, as few as three loci can provide precise estimates (compare Sim 2 to 7). When two sire contribute equally to a multiply mated brood, only three loci with five equally common alleles are

required to provide precise estimates of the frequency of multiple mating. Finally, precision also increased as the number of sires contributing to a multiply mated brood increased or as the skew in their paternity decreased (compare Sim 4 to 5, Sim 8 to 5, and Sim 9 to 2).

### Mutations

When unaccounted for in our model, mutations resulted in an overestimation of the frequency of multiple mating (Figure 3). Overall, the bias increased with mutation rate (compare Sim 12, 14, and 16). Interestingly however, mutations affected the two genetic criteria differently. For the first criterion (proportion of detected multiply mated broods) and for the second (number of haplotypes) without recombination, the bias increased with the number of loci used, while for the second criterion with recombination, the bias decreased (compare Sim 14 to 15 and 16 to 17). For example, at a mutation rate of 0.001 and when three loci were used, the bias in the estimate based on the first genetic criterion was less than 1% (0.004/0.50); while when eight loci were used, it was approximately 12% (0.062/0.50) (see Sim 16, 17). For the second criterion with recombination, when three loci were used,

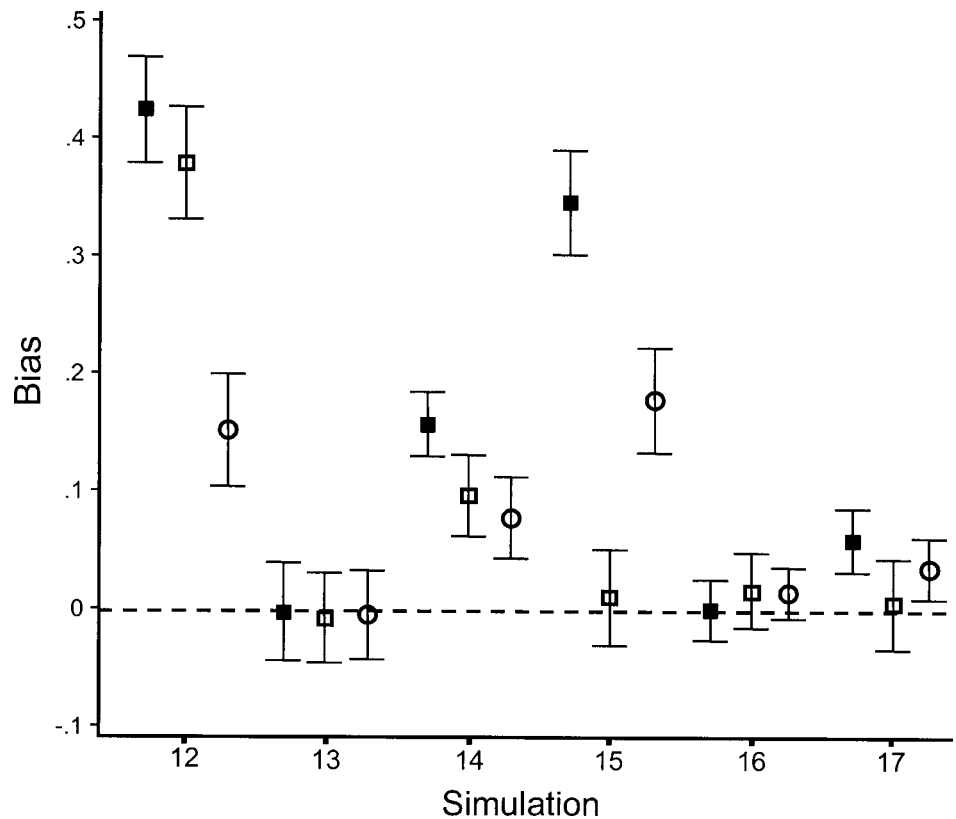
the bias was 6% (0.029/0.50); while when eight loci were used, it was 2% (0.009/0.50). Generally, the second genetic criterion appeared to be less sensitive to mutations, generating estimates that were less biased. When the correct mutation rate was included in the calculation of  $f_{mm}$ , no bias existed (compare Sim 12 to 13). However, mutations did decrease the precision of the estimates (compare Figure 3, Sim 13, and Figure 2, Sim 2). At lower mutation rates that are typical of microsatellite loci ( $\leq 0.001$ ), and when the mutation rate was included in the model, both criteria produced unbiased estimates, but the first criterion produced estimates that were more precise (data not shown).

### The Prior Probability $\Pr(f_{mm})$

Only when the correct prior probability distribution was used did the analysis provide accurate estimates of the frequency of multiple mating. If the prior probability was assumed to follow the uniform distribution when in fact it followed a normal distribution centered on a mean of 10% or 90%, the estimates of the frequency of multiple mating were overestimated or underestimated by about 8%, respectively. The uniform prior probability distribution assumed that the frequency of multiple mating was just as likely to be any value between 0% and 100%. Thus, when the prior probability actually followed the normal distribution centered on a mean of 10%, for example, the assumed uniform distribution underestimated the prior probability of lower frequencies of multiple mating, while overestimating higher frequencies.

### Multiple Mating and Kemp's Ridley Sea Turtles

In total, we analyzed data from 26 clutches from Kichler et al. (1999). When all clutches were analyzed simultaneously, we determined that the frequency of multiple mating was, on average, between 70% and 81% (Table 4), depending on the number of sires contributing to a multiply mated clutch and their reproductive skew. Generally, the estimated frequency of multiple mating decreased as the number of sires contributing to each multiply mated clutch increased or as the skew in their success decreased (Table 4). Thus, assuming that two sires did not have a greater reproductive skew than 66.7% and 33.3% (our Scenario 2), the estimate of 81% (95% CI: 58%–97%) can be considered as



**Figure 3.** Simulation results for the mutation analysis. The mean bias and 95% confidence are shown for simulations 12–17 (Table 3). The results from the first genetic criterion (proportion of broods with three or more paternal alleles) are on the left (closed square) of each simulation trio, those for the second criterion (number of haplotypes) with full recombination are in the middle (open square), and those for the second criterion without recombination are on the right (open circle).

a maximum. Alternatively, assuming that there were not more than three sires contributing equally (our Scenario 3), 70% (95% CI: 48%–88%) can be considered a minimum.

Under the four scenarios considered here, small clutches had a significantly lower frequency of multiple mating as compared to large clutches (Table 4). The frequency of multiple mating in small and large clutches ranged, on average, from 42% to 50% and from 89% to 92%, respectively, based on the number of sires (and their reproductive skew) contributing to a multiple mating. Because our model incorporates the number of offspring analyzed from each clutch, this difference cannot be attributed to the reduced ability of the genetic analysis to detect a multiple mating in small clutches (although the estimates of the frequency of multiple mating in small clutches were less precise).

### Discussion

In this paper we define the frequency of multiple mating ( $f_{mm}$ ) as the proportion

of broods within a sample that are multiply mated. We develop a single-sex model based on Bayes' rule for calculating the frequency of multiple mating. This model can be used when there is multiple mating by only one sex (defined at the level of the brood). Thus, it can estimate the frequency with which females mate with more than one male (genetic polyandry) or with which males mate with more than one female (genetic polygyny). We considered two genetic criteria for calculating  $f_{mm}$ : (1) the proportion of broods in which there are three or more paternal (or maternal) alleles detected at any one locus, and (2) the number of haplotypes (assuming full recombination or no recombination) detected in each brood. Although the second criterion has been shown to be more sensitive at detecting the actual number of mates contributing to multiply mated broods (DeWoody et al. 2000a,b), we found that it was generally inferior to the first criterion for calculating  $f_{mm}$ . This suggests that the frequency of multiple mating correlates better with the observed proportion of multiply mated

**Table 4. Estimates of the frequency of multiple mating in a population of Kemp's Ridley sea turtles**

Sires	Skew <sup>a</sup>	Frequency of multiple mating (%)			P
		All	Large	Small	
2	Even	78.1 (55–95)	91.2 (70–99)	48.8 (37–80)	.014
2	Skew	80.9 (58–97)	91.7 (71–99)	50.3 (20–83)	.017
3	Even	70.2 (48–88)	89.3 (67–99)	41.8 (14–73)	.007
3	Skew	72.0 (50–91)	90.9 (69–99)	42.7 (16–75)	.007

Four scenarios were considered for the number of sires and their reproductive skew. The frequency of multiple mating for all clutches ( $n = 26$ ), large clutches ( $n = 13$ ), and small clutches ( $n = 13$ ) are presented, and the 95% confidence intervals appear in parentheses. The  $P$ -value from an exact test (Equation 7) compares the frequency of multiple mating in large versus small clutches.

<sup>a</sup> Even denotes equal breeding success for each male and skew denotes each subsequent male has 50% as much the breeding success as the former male (see text).

broods (i.e., those containing three or more paternal or maternal alleles) than it does with the number of haplotypes observed in each brood. For the second criterion, we did not consider the proportion of broods with more than  $2^L$  haplotypes, because multiply mated broods rarely exceed this value when more than a few loci are used. However, when there is no recombination only three haplotypes are required to detect a multiply mated brood regardless of the number of loci used. In this case the loci can be treated as a single locus with each haplotype representing an “allele,” and the first approach can be used.

The probability that a multiple mating is detected depends on the number of individuals genetically contributing to a multiply mated brood as well as their skew in fertilization success. For example, when only two males contribute to a multiple mated brood and these males have very skewed paternities (95% and 5%), a multiply mated brood can be difficult to detect, particularly when only a few offspring are analyzed and the brood is very large. To be 80% confident that at least one offspring is sampled from the male having only 5% paternity, 31 offspring would have to be analyzed from each brood (Fiumera et al. 2001). To be 95% confident, 58 offspring would have to be analyzed. Thus, in systems with large skew in fertilization success, estimating the frequency of multiple mating with any precision will require large sample sizes. If the skew is assumed (in the model) incorrectly to be much lower, the frequency of multiple mating will be underestimated. Conversely, when the skew is assumed to be higher, the frequency will be overestimated. When the actual skew is unknown, it might be useful to estimate it by analyzing a subset of broods that, through behavioral observations, for example, are known to be multiple mated. While

the number of sires and their skew is generally out of the control of researchers, all else being equal, populations that have multiply mated broods consisting of genes from a greater number of individuals with less skew will have more precise estimates of the frequency of multiple mating.

Mutations result in an overestimation of the actual frequency of multiple mating (Figure 3). Interestingly, at mutation rates typical of microsatellite loci ( $\leq 0.001$ ), mutations have little effect on the accuracy of the estimates when only a small number of loci are used (e.g., three or fewer). However, when a large number of loci are used (eight or more), and when the first genetic criterion is used, the bias can be significant (e.g.,  $>12\%$ ; Figure 3). Thus, while increasing the number of loci can increase the precision of the estimates, it also increases the probability of a mutation, which decreases accuracy. If the mutation rate is unknown, it might be more conservative to use a smaller number of loci (e.g., three) to ensure that the bias in the estimates is small, or to use the second genetic criterion. However, if the mutation rate is known, it can be easily incorporated into the model, and unbiased estimates of the frequency of multiple mating can be obtained with any number of loci. Furthermore, in this case the first genetic criterion generates more precise estimates as compared to the second.

Most previous models that can be used to calculate the frequency of multiple mating have intrinsically assumed a uniform prior probability distribution. Thus, these models assume that the frequency of multiple mating is just as likely to be anywhere between 0 and 100%. In the context of parentage analysis, Neff et al. (2001) show that this may be the best assumption in the absence of other biological data. However, here we show

that assuming a uniform prior probability distribution can lead to a significant bias in the estimate of  $f_{mm}$ . Thus, if possible, other biological data should be used to estimate the prior probability. These data may include behavioral observations during mating (e.g., the frequency with which females are observed to mate with multiple males), or previous, independent genetic analysis (Neff et al. 2001). Alternatively, if sufficient samples and loci are used, such that all multiple matings are detected unambiguously, then the prior probability has little influence on the calculation of  $f_{mm}$ . This latter approach, however, can be costly and time consuming.

We have demonstrated our model with an empirical example involving endangered Kemp's Ridley sea turtles. Kichler et al. (1999) first observed that large clutches tended to have a greater frequency of three or more paternal alleles, as compared to small ones. However, they were unable to factor out the increased probability of detecting multiple matings in larger clutches (because a greater number of offspring were analyzed). Here, we used our model to calculate the frequency and 95% confidence interval of multiple mating in all of the clutches analyzed by Kichler et al., as well as in the small ( $<160$  eggs) and large ( $\geq 160$  eggs) clutches separately. Our model shows that the expected frequency of multiple mating in Kemp's Ridley sea turtles across all clutches was, on average, between 70% and 81%, depending on the assumed number of sires and their reproductive skew. As the number of sires that contribute to a multiple mating increased, or as the skew in their paternity decreased, the estimated frequency decreased (Table 4). Assuming that two males contribute to each multiple mating with skewed paternities of 66% and 33% [similar to the estimates of Kichler et al. (1999); also see Pearse et al. (2002)], the frequency of multiple mating across all clutches was 81% with a 95% confidence interval of 58%–97%. This is consistent with the estimates derived by Kichler et al. (1999), which were 70%–100% with a most likely estimate of 100%. Thus, multiple mating is quite high in Kemp's Ridley sea turtles and is comparably higher than in a recent study on painted turtles (*Chrysemys picta*) that reported only 30% of clutches were multiply mated (Pearse et al. 2002). Within small clutches of Kemp's Ridley (under the 66%:33% paternity skew scenario), the frequency of



multiple mating was 50% (20%–83%), and within large clutches it was significantly higher at 92% (71%–99%). Thus, females that lay larger clutches [likely larger females (Hirth 1980; Trexler et al. 1997)] are considerably more likely to have multiply mated as compared to females that lay smaller clutches. Pearse et al. (2002) also report a similar finding in their painted turtles, suggesting that increased multiple mating by large females may be common in turtles.

In conclusion, we have developed a model based on Bayes' rule to calculate the frequency of multiple mating from limited genetic data. The model permits the calculation of the associated confidence intervals and can include mutation and typing error rates. The model explicitly defines the prior probability of  $f_{mm}$  and shows that only when this probability distribution is known are accurate estimates obtained. We have demonstrated the model using genetic data from clutches of Kemp's Ridley sea turtles. This analysis shows that females that lay larger clutches are more likely to have multiply mated. The model is available as an executable program written in the C++ programming language at the Web site <http://publish.uwo.ca/~bneff/software.htm#FMM>. It should prove valuable to researchers assessing the frequency of multiple mating in nature.

## References

Barry FE, Weatherhead PJ, and Philipp DP, 1992. Multiple paternity in a wild population of northern water snakes, *Nerodia sipedon*. *Behav Ecol Sociobiol* 30:193–199.

Birdsall DA and Nash D, 1973. Occurrence of successful multiple insemination of females in natural populations of deer mice (*Peromyscus maniculatus*). *Evolution* 27:106–110.

DeWoody JA, DeWoody YD, Fiumera AC, and Avise JC, 2000a. On the number of reproductives contributing to a half-sib progeny array. *Genet Res Camb* 75:95–105.

DeWoody JA, Walker D, and Avise JC, 2000b. Genetic parentage in large half-sib clutches: theoretical estimates and empirical appraisals. *Genetics* 154:1907–1912.

Fiumera AC, DeWoody YD, DeWoody JA, Asmussen MA, and Avise JC, 2001. Accuracy and precision of methods to estimate the number of parents contributing to a half-sib progeny array. *J Hered* 92:120–126.

Hirth HF, 1980. Some aspects of the nesting behavior and reproductive biology of sea turtles. *Amer Zool* 20:507–523.

Jarne P and Lagoda JL, 1996. Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 8:285–288.

Jennions MD and Petrie M, 2000. Why do females mate multiply? A review of the genetic benefits. *Biol Rev* 75:21–64.

Kelly CD, Godin J-GJ, and Wright JM, 1999. Geographic variation in multiple paternity within natural populations of the guppy (*Poecilia reticulata*). *Proc Roy Soc Lond B* 266:2403–2408.

Kichler K, Holder MT, Davis SK, Márquez-M R, and Owens DW, 1999. Detection of multiple paternity in the Kemp's Ridley sea turtle with limited sampling. *Mol Ecol* 8:819–830.

Krebs JR and Davies NB, eds, 1997. *Behavioural ecology: an evolutionary approach*. Oxford: Blackwell Scientific.

Lewis PO, 2001. Phylogenetic systematics turns over a new leaf. *Trends Ecol Evol* 16:30–37.

Manly BFJ, 1997. *Randomization, bootstrapping and Monte Carlo methods in biology*, 2nd ed. London: Chapman and Hall.

Mendonça MT and Pritchard PCH, 1986. Offshore movements of post-nesting Kemp's Ridley sea turtles (*Lepidochelys kempi*). *Herpetologica* 42:373–381.

Milkman R and Zeitler RR, 1974. Concurrent multiple paternity in natural and laboratory populations of *Drosophila melanogaster*. *Genetics* 78:1191–1193.

Neff BD and Pitcher TE, 2002. Assessing the statistical power of genetic analyses to detect multiple mating in fish. *J Fish Biol* 61:739–750.

Neff BD, Repka J, and Gross MR, 2000a. Parentage analysis with incomplete sampling of candidate parents and offspring. *Mol Ecol* 9:515–528.

Neff BD, Repka J, and Gross MR, 2000b. Statistical confidence in models of parentage analysis with incomplete sampling: how many loci and offspring are needed? *Mol Ecol* 9:529–539.

Neff BD, Repka J, and Gross MR, 2001. A Bayesian framework for parentage analysis: the value of genetic and other biological data. *Theor Popul Biol* 59:315–331.

Ochando MD, Reyes A, and Ayala FJ, 1996. Multiple paternity in two natural populations (orchard and vineyard) of *Drosophila*. *Proc Natl Acad Sci USA* 93:11769–11773.

Pearse DE, Janzen FJ, and Avise JC, 2002. Multiple paternity, sperm storage, and reproductive success of female and male painted turtles (*Chrysemys picta*) in nature. *Behav Ecol Sociobiol* 51:164–171.

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

Reynolds JD, 1996. Animal breeding systems. *Trends Ecol Evol* 11:68–72.

Schwartz JM, McCracken GF, and Burghardt GM, 1989. Multiple paternity in wild populations of the garter snake, *Thamnophis sirtalis*. *Behav Ecol Sociobiol* 25:269–273.

Trexler JC, Travis J, and Dinep A, 1997. Variation among populations of the sailfin molly in the rate of concurrent multiple paternity and its implications for mating-system evolution. *Behav Ecol Sociobiol* 40:297–305.

Williams CJ and Everts S, 1989. The estimation of concurrent multiple paternity probabilities in natural populations. *Theor Popul Biol* 35:90–112.

Zane L, Nelson WS, Jone AG, and Avise JC, 1999. Microsatellite assessment of multiple paternity in natural populations of a live bearing fish, *Gambusia holbrooki*. *J Evol Biol* 12:61–69.

Received December 13, 2001

Accepted November 15, 2002

Corresponding Editor: Stephen J. O'Brien