Assessing the statistical power of genetic analyses to detect multiple mating in fishes

B. D. NEFF*‡ AND T. E. PITCHER†

Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada and †Department of Zoology, University of Toronto, 25 Harbord St, Toronto, Ontario M5S 3G5, Canada

A single-sex model is presented that calculates the probability of detecting multiple mating (*PrDM*) given genetic data from the single genetic parent and a sample of its offspring. The model incorporates the effects of numbers of loci, alleles, offspring and genetic parents contributing to the multiple mating, all of which effect *PrDM*. The model is used to determine the actual number of loci and offspring that are required to detect multiply mated broods with high probability (80 and 95%). For example, if two sires contribute with equal fertilization success to multiply mated broods, then only 10 offspring and one locus with seven equally common alleles are required to ensure that 80% of multiple mated broods are detected. Ninety-five per cent of multiple mated broods can be detected with 10 offspring and five loci with four equally common alleles. The utility of the model is demonstrated with biological examples addressing geographic variation in multiple paternity among natural populations of guppies *Poecilia reticulata* and mosquitofish *Gambusia holbrooki*.

© 2002 The Fisheries Society of the British Isles. Published by Elsevier Science Ltd. All rights reserved.

Key words: multiple mating; statistical power; paternity; reproductive skew; genetic markers; *Poeciliidae*.

INTRODUCTION

The recent discovery that multiple mating is prevalent in the animal kingdom has revolutionized the study of mating systems (Reynolds, 1996; Krebs & Davies, 1997; Jennions & Petrie, 2000). Multiple mating occurs when individuals of one sex mate with more than one individual of the opposite sex (Reynolds, 1996; Neff *et al.*, 2000*a*). Multiple mating can affect the intensity of natural and sexual selection (Andersson, 1994; Fleming & Gross, 1994; Evans & Magurran, 1999*a*, 2001), effective population sizes (Sugg & Chesser, 1994; Martinez *et al.*, 2000), and genetic variability and introgression (Moran & Garcia-Vazquez, 1998; Baer & Schmid-Hempel, 1999). In this paper the frequency of multiple mating is defined as the proportion of broods in the population that contain genes from at least two males (or alternatively two females). The term mating is used to reflect fertilization and not just copulation. Frequency differs from the degree of multiple mating, which instead calculates the number of individuals genetically contributing to each brood. The frequency of multiple mating is critical to understanding the evolution of mating systems and to the conservation of endangered populations (Kelly *et al.*, 1999; Kichler *et al.*, 1999; Zane *et al.*, 1999; Moore & Ball, 2002).

‡Author to whom correspondence should be addressed. Tel.: +1 519 850-2532; fax: +1 519 661-2014; email: bneff@uwo.ca

Generally, the frequency of multiple mating can be inferred from the degree of multiple mating, for example from the proportion of broods with two or more sires, and several models have been developed to calculate degree (Levine *et al.*, 1980; Pedersen & Boomsma, 1999; DeWoody *et al.*, 2000*a*,*b*; Lexer *et al.*, 2000). Large numbers of loci, however, can be required to accurately calculate the likely number of sires per brood. Alternatively, the frequency can be more efficiently calculated by analysing all broods simultaneously and adjusting the observed proportion of multiply mated broods (i.e. broods having three or more paternal alleles at a locus) by a correction factor (Milkman & Zeitler, 1974; Cobbs, 1977; Williams & Evarts, 1989; Kichler *et al.*, 1999). The correction factor accounts for the possibility that less than three paternal alleles are detected in a multiply mated brood, which can occur when two sires share a common genotype at the loci used in the analysis or when a sire's offspring are not sampled from the brood (due to random sampling error). The correction factor thus represents the probability of detecting a multiple mating (*PrDM*). *PrDM* depends on four parameters: (1) the number of loci; (2) the number of alleles (and their frequency); (3) the number of offspring analysed from the brood; (4) the number of genetic parents (and their reproductive skew). No model yet exists, however, that simultaneously accounts for all of these factors when calculating *PrDM*. The aim of the present study was to develop a model to calculate *PrDM* that incorporated any combination of these four parameters.

In the model, mating systems with single-sex multiple mating are considered. Single-sex multiple mating occurs when there is multiple mating by only one sex. Here, multiple mating is defined at the level of the brood (Neff *et al.*, 2000*a*). Therefore, single-sex multiple mating gives rise to broods that contain the genes of either a single female that has mated with multiple males or a single male that has mated with multiple females. As such, all of the young within a brood are either full- or half-sibs. The model requires single-locus genetic data, such as microsatellites or allozymes, from the known genetic parent and a sample of the brood. A genetic sample from the breeding population is also required to estimate allele frequencies.

The model is tested with two biological examples based on published data. The first example addresses geographic variation in multiple paternity among wild populations of guppies *Poecilia reticulata* (Peters) (Kelly *et al.*, 1999). Guppies experience two distinct predation regimes. Guppies are considered to be in high predation locales when they co-exist with the piscivorous cichlid *Crenicichla alta* Eigenmann and low predation locales when they co-exist with the gape-limited cyprinodontid *Rivulus hartii* (Boulenger), which has a limited impact only on juvenile guppies (Farr, 1975; Houde, 1997). As the costs of mating increases due to increased predation, female guppies are less selective in their choice of mates (Godin & Briggs, 1996; Gong & Gibson, 1996) and males alter their mating tactics to include fewer conspicuous courtship displays and more sneak copulations (Luyten & Liley, 1985; Houde & Endler, 1990). Therefore, it might be expected that multiply sired broods are more common in high predation populations (Kelly *et al.*, 1999; Houde, 1988; Evans & Magurran, 1999*a*, *b*, 2000). The second example calculates and compares the frequency of multiple mating in two wild populations of the related mosquitofish *Gambusia holbrooki* Girard from Fisher Pond and Fire Pond (Zane *et al.*, 1999). Previous analysis of

multiple mating in other populations of mosquitofish found that c. 55–65% of broods were multiply sired (Chesser et al., 1984; Greene & Brown, 1991). Genetic data from the populations studied by Zane et al. (1999), however, suggest that genetic variation is relatively low. Thus, if females mate multiply to increase genetic variation in their sperm supply (Jennions & Petrie, 2000), then the frequency of multiple mating should be higher (>65%) in Fisher Pond and Fire Pond.

MATERIAL AND METHODS

THE MODEL

The model builds on those developed by Cobbs (1977) and Kichler et al. (1999). It calculates the probability that a multiple mating is detected (given that there is a multiple mating) using a Monte Carlo simulation (Manly, 1997). While it might be desirable to generate a statistic directly, this is exceedingly difficult when multiple loci are used (Harshman & Clarke, 1998). The model requires four parameters: (1) the number of loci; (2) the number of alleles and their frequencies within the breeding population; (3) the number of offspring analysed from the brood; (4) the number of mates (e.g. sires) and their relative contributions (i.e. skew in fertilization success within the brood). The genotype of a known genetic parent can also be an input for individual broods. When estimating *PrDM* for a population or set of independent broods it may be omitted and instead randomly generated by the model based on the population allele frequency data. All of the parameters influence *PrDM* and must therefore be accounted for in the analysis (Cobbs, 1977). Because the number of mates and their reproductive skew will probably be unknown methods to estimate them are discussed below. Also, Harshman & Clark (1998) develop a model based on a Monte Carlo simulation to directly estimate reproductive skew from genetic data, assuming that the probability that a female remates follows a Poisson distribution and the skew follows a geometric distribution.

The simulation (Fig. 1) generates genotypes for a genetic mother (or father) and its m mates based on the genetic polymorphism data. Each of the b offspring within the brood is then generated by selecting one of the m mates according to their relative fertilization success (skew). For example, if two mates contribute to a brood with relative success of 75 and 25%, then it is three times more likely that the first mate will produce any particular offspring and on average the first mate will produce 75% of the offspring within a brood. One of the two alleles from each parent at each locus is randomly assigned (with equal probability) to the offspring. The genotypes of the b offspring are then analysed to determine the number of different alleles from the mates. The multiple mating is detected when more than two different alleles from the mates at any locus are identified in the b offspring (Birdsall & Nash, 1973; Barry et al., 1992; Baker et al., 1999). The multiple mating could also be detected when only two paternal alleles are detected, but the frequency of the two alleles in the offspring differ from the expected 1 : 1 ratio under Mendelian inheritance and assuming that there is only a single heterozygous sire (Barry et al., 1992). This latter possibility, however, is not considered because it adds little resolution to the analysis when the loci used have more than two common alleles (as is the case for most microsatellites) and when only a limited number of offspring are analysed (Ochando et al., 1996). In cases where a mate's allele is ambiguous, for example when the known genetic parent and offspring are heterozygous with the same alleles, the minimum number of alleles is calculated. The entire process is repeated for a total of 10 000 broods and *PrDM* is calculated as the proportion of these broods that a multiple mating (at least three paternal or maternal alleles) is detected.

MULTIPLE MATING IN GUPPIES

Kelly et al. (1999) used two microsatellite loci to estimate the proportion of broods with multiple paternity from 10 wild guppy populations (five populations from each of high and low predation locales). They reported the results of an analysis of covariance
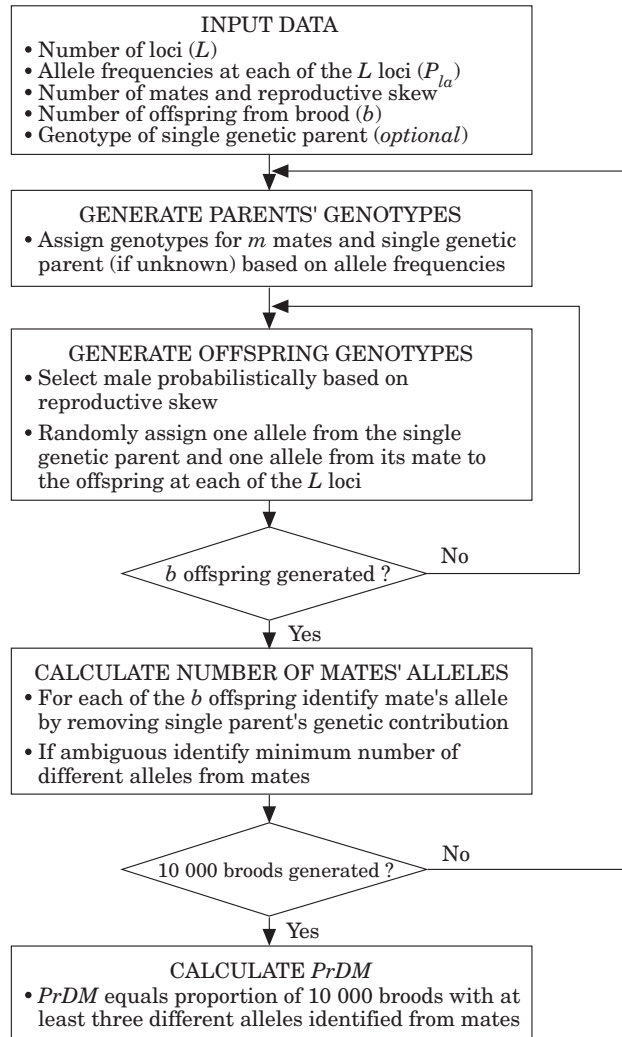
```
┌─────────────────────────────────────────────────────┐
│                    INPUT DATA                         │
│ • Number of loci (L)                                  │
│ • Allele frequencies at each of the L loci (P_la)     │
│ • Number of mates and reproductive skew               │
│ • Number of offspring from brood (b)                  │
│ • Genotype of single genetic parent (optional)        │
└─────────────────────────────────────────────────────┘
                          │
                          ▼
┌─────────────────────────────────────────────────────┐
│             GENERATE PARENTS' GENOTYPES               │
│ • Assign genotypes for m mates and single genetic     │
│   parent (if unknown) based on allele frequencies     │
└─────────────────────────────────────────────────────┘
                          │
                          ▼
┌─────────────────────────────────────────────────────┐
│            GENERATE OFFSPRING GENOTYPES               │
│ • Select male probabilistically based on              │
│   reproductive skew                                   │
│ • Randomly assign one allele from the single          │
│   genetic parent and one allele from its mate to      │
│   the offspring at each of the L loci                 │
└─────────────────────────────────────────────────────┘
                          │
                          ▼
              ◇ b offspring generated ? ◇ ──No──┐
                          │ Yes
                          ▼
┌─────────────────────────────────────────────────────┐
│          CALCULATE NUMBER OF MATES' ALLELES           │
│ • For each of the b offspring identify mate's allele  │
│   by removing single parent's genetic contribution    │
│ • If ambiguous identify minimum number of             │
│   different alleles from mates                        │
└─────────────────────────────────────────────────────┘
                          │
                          ▼
              ◇ 10 000 broods generated ? ◇ ──No──┐
                          │ Yes
                          ▼
┌─────────────────────────────────────────────────────┐
│                  CALCULATE PrDM                       │
│ • PrDM equals proportion of 10 000 broods with at     │
│   least three different alleles identified from mates │
└─────────────────────────────────────────────────────┘
```

FIG. 1. Schematic of the Monte Carlo simulation used to calculate the probability of detecting a multiple mating (*PrDM*).

(ANCOVA) test which controls only for differences in the number of offspring analysed per brood and showed that the predation regime significantly affects the frequency of multiple paternity in guppies ($F_{1,7}=8\cdot31$, $P=0\cdot024$). The loci, however, were more polymorphic in the high predation populations which will increase the chance of detecting multiple matings and could lead to spurious differences in the frequency of multiple mating among populations.

For each population *PrDM* was calculated using the mean and s.D. in the number of offspring analysed per brood reported in Kelly *et al.* (1999) and the genetic polymorphism data reported in Kelly (1999). Because the number of sires contributing to each brood was unknown six situations were considerd: (1) two sires with equal fertilization success (mean paternity of 50% for each male) for broods from all populations; (2) two sires with skewed success (66·7 and 33·3%) for all populations; (3) three sires with equal success (33·3% for each male) for broods from all populations; (4) three sires with skewed success (57, 28·5 and 14·5%) for all populations; (5) two sires with skewed success (66·7 and

33·3%) for populations from the high predation locales and three sires with skewed success (57, 28·5 and 14·5%) for populations from the low predation locales; (6) two sires with skewed success (66·7 and 33·3%) for populations from the low predation locales and three sires with skewed success (57, 28·5 and 14·5%) for populations from the high predation locales. The first four situations have no difference in the number of sires between populations from the two predation regimes, while the last two situations consider differing numbers. The six situations cover a variety of multiple matings and probably encompass the actual distribution of fertilization success in these populations (Houde, 1997; Kelly *et al.*, 1999). For each of the six situations ANCOVA was conducted with the observed frequency of multiply mated broods as the dependent variable, predation regime (low *v.* high) as the independent variable and *PrDM* as the covariate. Both the frequency and *PrDM* were first arcsine transformed to restore normality to the data (Zar, 1999).

MULTIPLE MATING IN MOSQUITOFISH

Zane *et al.* (1999) used three microsatellite loci to estimate the proportion of broods with multiple paternity from two wild mosquitofish populations, Fisher Pond and Fire Pond. The two ponds are located *c.* 10 km apart at the Savannah River site near Aiken, South Carolina, U.S.A. They are believed to be genetically distinct (Zane *et al.*, 1999). Zane *et al.* (1999) found that based on direct paternal allele counts in each brood the frequency of multiple mating in Fisher Pond was 84% and in Fire Pond it was 88%. Because the combined exclusion probability across the three loci was only 0·84 in each population, they concluded that their frequency estimates were conservative (i.e. minimal estimates). They were unable, however, to calculate the actual accuracy of their estimates.

For each population *PrDM* was calculated using the mean number of offspring and the genetic data reported in Zane *et al.* (1999). Here two situations were considered for the number of sires contributing to multiply mated broods: (1) two sires with equal fertilization success (50% each); (2) two sires with skewed success (66·7 and 33·3%). For each of the two situations, *PrDM* was used to adjust the observed frequency of multiple mating according to the following calculation: actual frequency of multiple mating=(observed frequency of multiple mating)$(PrDM)^{-1}$.

RESULTS

THE MODEL

Increasing the numbers of loci, alleles, offspring and genetic parents increases *PrDM* (Fig. 2 and Table I). Increasing either the number of loci used to genotype the individuals or using loci with more alleles or with frequencies that are less skewed (i.e. more effective alleles; Neff *et al.* 2000*b*) increases genetic polymorphism. In turn, it is more likely that each genetic parent is heterozygous and contains different alleles and therefore *PrDM* increases. Loci with less skewed allele frequencies are considerably more valuable than loci with highly skewed frequencies. For example, when allele frequencies are skewed, nearly twice as many alleles are required to obtain equivalent *PrDM* values as loci with even allele frequencies (Table I).

It is more likely that a multiple mating is detected when the known parent is homozygous (unpubl. data). *PrDM*, however, increases with locus polymorphism (number of alleles) even though polymorphism decreases the probability that the known parent is homozygous. This is attributed to the predominant positive effect of polymorphism on the probability that the known parent's mates are heterozygous and possess different alleles, which are more likely to be detected in the sample of offspring.
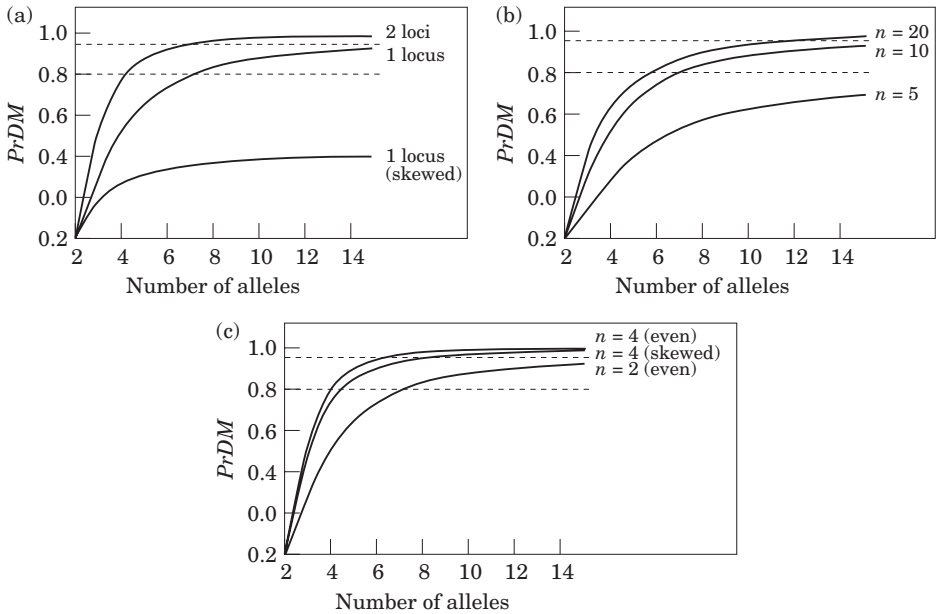
FIG. 2. The effects of the numbers of loci, alleles, offspring and genetic parents on the probability of
detecting a multiple mating (PrDM). (a) Genetic polymorphism. PrDM increases with the number
of loci, the number of alleles and as the frequency of the alleles become less skewed. Skewed
frequencies were calculated such that each subsequent allele had 50% of the frequency of the
previous allele (see Table I). (b) Number of offspring. PrDM increases with the number of offspring
analysed from the brood. (c) Number of sires. PrDM increases with the number of parents
genetically contributing to the brood and as the skew in fertilization success becomes less skewed.
Skewed fertilization success was calculated analogously to the skewed allele frequencies. For each
situation, where not indicated, it was assumed that 10 offspring were analysed, one locus was
used with even allele frequencies, and two males contributed equally to the multiply mated brood.
– – –, PrDM values of 80 and 95%.

Increasing the number of offspring analysed from a brood increases the
probability that the alleles from each genetic parent are sampled and therefore
increases *PrDM* [Fig. 2(b)]. Finally, increasing the number of parents that
genetically contribute to the brood, or equalizing the relative contribution of
each parent, increases the probability that more than two different alleles are
detected within the brood and therefore increases *PrDM* [Fig. 2(c)]. Generally,
however, the number of sires contributing to a brood and their reproductive
skew will be a property of the mating system and out of the hands of the
researcher. Nevertheless, all else equal, mating systems with more sires and lower
skew will have greater statistical power to detect multiple matings.

While it requires relatively few loci, alleles and offspring to obtain *PrDM*
values of 80%, it requires considerably more to obtain *PrDM* values of 95%. For
example, assuming that two sires contribute to multiply mated broods with equal
success, a *PrDM* value of 80% can be obtained by analysing 10 offspring with
one locus with seven equally common alleles, or two loci with four equally
common alleles [Fig. 2(a)]. To obtain a *PrDM* value of 95%, 10 offspring would
have to be analysed with one locus with 15 equally common alleles or two loci
with seven equally common alleles. Thus, loci with approximately twice as many

TABLE I. *PrDM* values for combinations of numbers of loci, offspring, and alleles. Multiply mated broods were assumed to be the product of either two sires with fertilization success of two-thirds and one-third (first *PrDM* number) or three sires with equal success (one-third each; second number)

Loci	Offspring	Alleles									
		Even					Skewed*				
		3	4	5	10	20	3	4	5	10	20
1	5	0·13	0·26	0·36	0·57	0·66	0·09	0·15	0·18	0·21	0·21
		0·21	0·42	0·56	0·78	0·87	0·14	0·24	0·29	0·33	0·34
	10	0·28	0·49	0·62	0·84	0·92	0·19	0·29	0·33	0·38	0·39
		0·47	0·74	0·85	0·97	0·99	0·32	0·47	0·53	0·59	0·59
	20	0·40	0·61	0·73	0·92	0·97	0·25	0·37	0·42	0·47	0·47
		0·66	0·87	0·94	0·99	1·00	0·45	0·60	0·66	0·71	0·71
2	5	0·24	0·44	0·57	0·76	0·82	0·17	0·27	0·32	0·37	0·37
		0·38	0·66	0·78	0·94	0·97	0·27	0·42	0·50	0·55	0·55
	10	0·48	0·74	0·85	0·96	0·98	0·33	0·49	0·56	0·61	0·61
		0·72	0·93	0·97	1·00	1·00	0·54	0·72	0·77	0·82	0·83
	20	0·63	0·85	0·93	0·99	1·00	0·44	0·60	0·66	0·72	0·72
		0·88	0·98	0·99	1·00	1·00	0·70	0·84	0·88	0·91	0·92
5	5	0·47	0·72	0·80	0·86	0·86	0·36	0·53	0·59	0·65	0·65
		0·68	0·91	0·96	0·98	0·99	0·54	0·74	0·80	0·85	0·85
	10	0·80	0·95	0·97	0·98	0·98	0·64	0·80	0·86	0·89	0·89
		0·95	1·00	1·00	1·00	1·00	0·86	0·95	0·98	0·99	0·99
	20	0·92	0·99	1·00	1·00	1·00	0·77	0·90	0·93	0·96	0·96
		0·99	1·0	1·00	1·00	1·00	0·95	0·99	0·99	1·00	1·00

*Skewed allele frequencies were calculated such that each subsequent allele had 50% the frequency of the previous allele. As an example, for three alleles the frequencies would be 0·571, 0·286 and 0·143.

alleles are required to raise *PrDM* from 80 to 95%. Nevertheless, with microsatellite genetic markers, *PrDM* values of 95% should be obtainable.

MULTIPLE MATING IN GUPPIES

The mean *PrDM* for the low predation populations ranged from 0·31–0·45 and for high predation populations it ranged from 0·62–0·80 depending on the reproductive skew situation. Thus, the probability of detecting a multiple mating in low predation populations was considerably lower than in the high predation populations. Subsequently, it was found that *PrDM* had a highly significant effect and explained most of the variation in the observed frequency of multiple mating among the populations (80–99%; Table II). The significance of predation intensity, however, was dependent on the number of sires. Provided that the low predation populations had the same or more sires contributing to each multiply mated brood as compared to the high predation populations, there was a small but significant effect of predation on the frequency of multiple mating. If the high predation populations had more sires contributing to each multiply mated brood, however, then there was no significant effect of predation (Table II).

TABLE II. Analysis of variance of the frequency of multiple mating in guppies between high and low predation regimes. Six possible situations were considered for the number of sires and their relative paternity (equal or skewed) within the broods. The proportion of the explained variation, per cent in parentheses, was partitioned between Predation and the covariate $PrDM$

Number of sires		Source	r^2	P
Low predation ($n=5$)	High predation ($n=5$)			
1. Two equal	Two equal	Model	0·92	<0·001
		Predation	(7)	0·047
		$PrDM$	(93)	<0·001
2. Two skewed	Two skewed	Model	0·92	<0·001
		Predation	(7)	0·047
		$PrDM$	(93)	<0·001
3. Three equal	Three equal	Model	0·92	<0·001
		Predation	(7)	0·055
		$PrDM$	(93)	<0·001
4. Three skewed	Three skewed	Model	0·92	<0·001
		Predation	(6)	0·054
		$PrDM$	(94)	<0·001
5. Three skewed	Two skewed	Model	0·92	>0·001
		Predation	(20)	0·005
		$PrDM$	(80)	<0·001
6. Two skewed	Three skewed	Model	0·93	<0·001
		Predation	(1)	0·395
		$PrDM$	(99)	<0·001

MULTIPLE MATING IN MOSQUITOFISH

If two sires contributed equally to multiply mated broods, then $PrDM$ in Fisher Pond and Fire Pond would be 0·975 and 0·982, respectively. If two sires contributed with skewed success, then $PrDM$ would be 0·967 and 0·974 in the two respective populations. Thus, with only three loci, Zane et al. (1999) had high statistical power to detect multiply mated broods. Based on the calculation of $PrDM$, the actual frequencies of multiple mating in Fisher Pond and Fire Pond were 86·2 and 89·6%, respectively, for two sires with equal success and 86·9 and 90·3% for two sires with skewed success. Thus, consistent with the conclusions of Zane et al. (1999), the frequency of multiple mating in the two populations is very similar.

DISCUSSION

In this paper a model is developed to calculate the statistical power of genetic analyses to detect multiple mating. The model incorporates the number of loci, number of alleles and their frequencies, number of offspring analysed from each brood, and number of genetic sires and their reproductive skew. The probability of detecting a multiple mating, and thus statistical power, increases with the number of loci and offspring analysed, and the number of genetic parents

contributing to the brood (Fig. 2). It decreases as the population frequency of the alleles carried by the single genetic parent increases (and if the parent is heterozygous) and as reproductive skew among the parents increases (Fig. 2). Analogous results have been found in the context of genetic parentage analysis and population assignments (Bernatchez & Duchesne, 2000; Neff *et al.*, 2000*b*).

Generally, only a few loci and samples are required to obtain high statistical power to detect multiple matings. For example, when two sires contribute with equal success to a brood, only 10 offspring (per brood) and one locus with seven equally common alleles are required to ensure that 80% of multiply mated broods are detected. *PrDM* values of 95% can be obtained with 10 offspring and two loci with seven equally common alleles or five loci with four equally common alleles. When allele frequencies are skewed, considerably more loci are required to obtain equivalent statistical power (Table I). Thus, future studies may be better to focus on loci with fewer, evenly distributed alleles. Nevertheless, with genetic markers like microsatellites, high statistical power should be feasible in many studies.

The effectiveness of the model has been demonstrated by analysing the effect of predation intensity on the frequency of multiple mating among wild guppy populations (Kelly *et al.*, 1999). There were two important consequences of *PrDM* on the analysis. First, there was a higher chance of detecting multiple matings in the high predation populations because of the increased genetic polymorphism (i.e. number of alleles at each locus) in these populations. Most (80–99%) of the explained variation in the frequency of multiple mating was attributed to differences in *PrDM* (Table II). As such, the observed difference in frequencies of multiple mating between high (64·2% of broods) and low (24·6% of broods) predation populations observed by Kelly *et al.* (1999) may be attributed to both biological and statistical effects. Second, the frequency of multiple mating was significantly higher in the high predation populations as compared to the low predation populations, unless the number of sires contributing to a multiply mated brood was greater in the high predation populations. In this latter case there would be a difference in the degree of multiple mating, but not the frequency. Both results are interesting and biologically important.

In the second example on mosquitofish, it has been shown that with only three microsatellite loci Zane *et al.* (1999) had high statistical power to detect multiple matings in their two wild populations. Assuming that only two males contributed to multiply mated broods with equal or marginally skewed fertilization success, >96% of such broods should have been detected. Interestingly, *PrDM* was considerably higher than the parentage exclusion probability (=84%). Thus, exclusion probabilities do not accurately reflect the probability of detecting multiple matings. Incorporating *PrDM* into the observed frequency of multiple mating, it was concluded that the actual frequency of multiple mating was very similar in the two populations (Fisher Pond and Fire Pond), ranging between 86 and 90%. These estimates are substantially higher than estimates of 55–65% from other populations. Because the Fisher Pond and Fire Pond populations have lower genetic variability, the increased rate of multiple mating in these populations might reflect females attempting to increase the genetic diversity in their sperm supplies.

In conclusion, this study highlights a general point for studies assessing the frequency of multiple mating: *PrDM* directly affects the proportion of multiple matings that are detected and therefore should be included in statistical analyses. The accuracy of statistical analyses can be maximized by ensuring that *PrDM* is not only as high as possible, but also equal between groups that are compared (e.g. the high and low predation populations in guppies or the two populations of mosquitofish). Equating *PrDM* between groups may involve increasing the number of loci for groups that have low genetic polymorphism or increasing the number of offspring analysed from each brood. The model can be used with polymorphism data collected from a preliminary screening of the groups to determine the optimal approach to achieve a desired *PrDM*. It is also possible that the Monte Carlo simulation developed here could be incorporated into a Bayesian framework to provide estimates of the frequency of multiple mating. The software is available as a downloadable executable file (http://publish.uwo.ca/~bneff/software.htm PrDM).

References

Andersson, M. (1994). *Sexual Selection*. Princeton: Princeton University Press.

Baer, B. & Schmid-Hempel, P. (1999). Experimental variation in polyandry affects parasite loads and fitness in a bumble-bee. *Nature* **397,** 151–154.

Baker, R. J., Makova, K. D. & Chesser, R. K. (1999). Microsatellites indicate a high frequency of multiple paternity in *Apodemus* (Rodentia). *Molecular Ecology* **8,** 107–111.

Barry, F. E., Weatherhead, P. J. & Philipp, D. P. (1992). Multiple paternity in a wild population of northern water snakes, *Nerodia sipedon*. *Behavioral Ecology and Sociobiology* **30,** 193–199.

Bernatchez, L. & Duchesne, P. (2000). Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences* **57,** 1–12.

Birdsall, D. A. & Nash, D. (1973). Occurrence of successful multiple insemination of females in natural populations of deer mice (*Peromysus maniculatus*). *Evolution* **27,** 106–110.

Chesser, R. K., Smith, M. W. & Smith, M. H. (1984). Biochemical genetics of mosquitofish III. Incidence and significance of multiple insemination. *Genetica* **64,** 77–81.

Cobbs, G. (1977). Multiple insemination and male sexual selection in natural populations of *Drosophila pseudoobscura*. *American Naturalist* **111,** 641–656.

DeWoody, J. A., DeWoody, Y. D., Fiumera, A. C. & Avise, J. C. (2000*a*). On the number of reproductives contributing to a half-sib progeny array. *Genetics Research (Cambridge)* **75,** 95–105.

DeWoody, J. A., Walker, D. & Avise, J. C. (2000*b*). Genetic parentage in large half-sib clutches: theoretical estimates and empirical appraisals. *Genetics* **154,** 1907–1912.

Evans, J. P. & Magurran, A. E. (1999*a*). Male mating behaviour and sperm production characteristics under varying sperm competition risk in guppies. *Animal Behaviour* **58,** 1001–1006.

Evans, J. P. & Magurran, A. E. (1999*b*). Geographic variation in sperm production by Trinidadian guppies. *Proceedings of the Royal Society of London (B)* **266,** 2083–2087.

Evans, J. P. & Magurran, A. E. (2000). Multiple benefits of multiple mating in guppies. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 10074–10076.

Evans, J. P. & Magurran, A. E. (2001). Patterns of sperm precedence and predictors of paternity in the Trinidadian guppy. *Proceedings of the Royal Society of London (B)* **268,** 719–724.

Farr, J. A. (1975). The role of predation in the evolution of social behavior of natural populations of the guppy, *Poecilia reticulata* (Pisces: Poeciliidae). *Evolution* **29,** 151–158.

Fleming, I. A. & Gross, M.R. (1994). Breeding competition in a pacific salmon (coho, *Oncorhynchus kisutch*)—measures of natural and sexual selection. *Evolution* **48,** 637–657.

Godin, J-G. J. & Briggs, S. E. (1996). Female mate choice under predation risk in the guppy. *Animal Behaviour* **51,** 117–130.

Gong, A. & Gibson, R. M. (1996). Reversal of female preference after visual exposure to a predator in the guppy, *Poecilia reticulata*. *Animal Behaviour* **52,** 1007–1015.

Greene, J. M. & Brown, K. L. (1991). Demographic and genetic characteristics of multiply inseminated female mosquitofish (*Gambusia affinis*). *Copeia* **1991,** 434–444.

Harshman, L. G. & Clarke, A. G. (1998). Inference of sperm competition from broods of field-caught Drosophila. *Evolution* **52,** 1334–1341.

Houde, A. E. (1988). The effects of female choice and male-male competition on the mating success of male guppies. *Animal Behaviour* **36,** 888–896.

Houde, A. E. (1997). *Sex, Color, and Mate Choice in Guppies*. Princeton: Princeton University Press.

Houde, A. E. & Endler, J. A. (1990). Correlated evolution of female mating preferences and male color patterns in the guppy, *Poecilia reticulata*. *Science* **248,** 1405–1408.

Jennions, M. D. & Petrie, M. (2000). Why do females mate multiply? A review of the genetic benefits. *Biological Reviews (Cambridge Philosophical Society)* **75,** 21–64.

Kelly, C. D. (1999). Effects of predation risk on the variance in male mating success in the Trinidadian guppy (*Poecilia reticulata*). (MSc Dissertation). Mt Allison University, Sackville NB, Canada.

Kelly, C. D., Godin, J-G. J. & Wright, J. M. (1999). Geographic variation in multiple paternity within natural populations of the guppy (*Poecilia reticulata*). *Proceedings of the Royal Society of London (B)* **266,** 2403–2408.

Kichler, K., Holder, M. T., Davis, S. K., Márquez, R. & Owens, D. W. (1999). Detection of multiple paternity in the Kemp's ridley sea turtle with limited sampling. *Molecular Ecology* **8,** 819–830.

Krebs, J. R. & Davies, N. B. (Eds) (1997). *Behavioural Ecology: An Evolutionary Approach*. Oxford: Blackwell Science.

Levine, L., Asmussen, M., Olvera, O., Powell, J. R., de la Rosa, M. E., Salceda, V. M. Gaso, M. I., Guzman, J. & Anderson, W. W. (1980). Population genetics of Mexican *Drosophila*. 5. A high-rate of multiple insemination in a natural population of *Drosophila pseudoobscura*. *American Naturalist* **116,** 493–503.

Lexer, C., Heinze, B., Gerber, S., Macalka-Kampfer, S., Steinkellner, H., Kremer, A. & Glössl, J. (2000). Microsatellite analysis of maternal half-sib families of *Quercus robur*, pedunculate oak: II. inferring the number of pollen donors from the offspring. *Theoretical Applications of Genetics* **100,** 858–865.

Luyten, P. H. & Liley, N. R. (1985). Geographic variation in the sexual behaviour of the guppy. *Behaviour* **95,** 164–179.

Manly, B. F. J. (1997). *Randomization, Bootstrapping and Monte Carlo Methods in Biology,* 2nd edn. London: Chapman & Hall.

Martinez, J. L., Moran, P., Perez, J., de Gaudemar, B., Beall, E. & Garcia-Vazquez, E. (2000). Multiple paternity increases effective population size of southern Atlantic salmon populations. *Molecular Ecology* **9,** 293–298.

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

Moore, M. K. & Ball, R. M. (2002). Multiple paternity in loggerhead turtle (*Caretta caretta*) nests on Melbourne Beach, Florida: a microsatellite analysis. *Molecular Ecology* **11,** 281–288.

Moran, P. & Garcia-Vazquez, E. (1998). Multiple paternity in Atlantic salmon: A way to maintain genetic variability in relicted populations. *Journal of Heredity* **89,** 551–553.

Neff, B. D., Repka, J. & Gross, M. R. (2000*a*). Parentage analysis with incomplete sampling of candidate parents and offspring. *Molecular Ecology* **9,** 515–528.

Neff, B. D., Repka, J. & Gross, M. R. (2000*b*). Statistical confidence in parentage analysis with incomplete sampling: how many loci and offspring are needed? *Molecular Ecology* **9,** 529–540.

Ochando, M. D., Reyes, A. & Ayala, F. J. (1996). Multiple paternity in two natural populations (orchard and vineyard) of *Drosophila. Proceedings of the National Academy of Sciences of the United States of America* **93,** 11769–11773.

Pedersen, J. S. & Boomsma, J. J. (1999). Multiple paternity in social Hymenoptera: estimating the effective mate number in single-double mating populations. *Molecular Ecology* **8,** 577–587.

Reynolds, J. D. (1996). Animal breeding systems. *Trends in Ecology and Evolution* **11,** 68–72.

Sugg, D. W. & Chesser, R. K. (1994). Effective population sizes with multiple paternity. *Genetics* **137,** 1147–1155.

Williams, C. J. & Evarts, S. (1989). The estimation of concurrent multiple paternity probabilities in natural populations. *Theoretical Population Biology* **35,** 90–112.

Zane, L., Nelson, W. S., Jones, A. G. & Avise, J. C. (1999). Microsatellite assessment of multiple paternity in natural populations of a live-bearing fish, *Gambusia holbrooki. Journal of Evolutionary Biology* **12,** 61–69.

Zar, J. H. (1999). *Biostatistical Analysis*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.