Behav Ecol Sociobiol (2017) 71: 12
DOI 10.1007/s00265-016-2240-x

METHODS PAPER

Statistical power to detect multiple paternity in populations of highly fertile species: how many females and how many offspring should be sampled?

David Veliz[1] · **Pierre Duchesne**[2] · **Noemi Rojas-Hernandez**[1] · **Luis M Pardo**[3,4]

Received: 13 February 2016 / Revised: 25 September 2016 / Accepted: 17 October 2016 / Published online: 3 December 2016

Abstract
One of the central issues of behavioral ecology focuses on the probability of detecting multiple paternity in a scenario of polygamy. The main problem for this kind of analysis arises in species with large number of offspring in the same litter and large population sizes in which only a small fraction of progeny and females can be analyzed. Here, we present a method to estimate the statistical power to detect multiple paternity for these species. Since calculations involved handling of very large numbers, Ramanujan's approximation to factorials was used to make them possible in the R software. We exemplified this method using features observed in crabs; (i) females carry thousands or millions of embryos per brood, (ii) typically less than 50% of females show multiple paternity, and (iii) high contribution of a single male (>90%) in a brood. Genetic parental analysis assumes the use of loci that allow maximal discrimination among individuals. The results showed that the number of females sampled is an important point to be considered to detect multiple paternity with high statistical power. Comparisons of different numbers of sampled females and embryos showed that 20 larvae from 50 females present satisfactory statistical power even when all males except the main contributor sired a modest number of embryos and only a small proportion of females showed embryos sired by more than one male. The proposed method can improve the sampling design in order to reach sufficient levels of statistical power when testing for multiple paternity in species with high fecundity, a common characteristic in both terrestrial and aquatic environments.

Significance statement
To detect multiple paternity in highly fertile species, researchers commonly use the probability of detecting multiple mating (PrDM). Although the PrDM is a powerful tool to detect multiple paternity in a litter/brood/clutch, this analysis neglects the estimation of the statistical power at the population level. Here, we developed an analytical method to assess the statistical power to detect multiple paternity considering brood size and the number of females sampled in a population. The model tested with 1,000,000 of embryos per brood (as in marine crabs), different numbers of embryos per brood, and different numbers of females analyzed reached values of statistical power greater than 99% when the first male sired 90% of the progeny and 50% of females had multiple paternity. This analysis showed the importance of focusing on the experimental unit in the experimental design in studies where multiple paternity is being tested.

Keywords Polyandry detection · Theory of sampling · Highly fecund species

Communicated by T. Breithaupt

Electronic supplementary material The online version of this article (doi:10.1007/s00265-016-2240-x) contains supplementary material, which is available to authorized users.

✉ David Veliz
dveliz@uchile.cl

[1] Departamento de Ciencias Ecológicas, Instituto de Ecología y Biodiversidad (IEB), Núcleo Milenio en Ecología y Manejo Sustentable (ESMOI), Universidad de Chile, Santiago, Chile

[2] Département de Biologie, Université Laval, Québec, Canada

[3] Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Valdivia, Chile

[4] Centro de Investigación de Dinámica de Ecosistemas Marinos de Altas Latitudes (IDEAL), Valdivia, Chile

Introduction

Detection of polygamy in natural populations is a core issue in behavioral and evolutionary ecology. The use of an accurate method to detect polyandry is therefore central to the assessment of whether sexual selection is strong or weak in many species. Sexual selection becomes strong when only a subset of male sires offspring. In contrast, sexual selection is weakened as the number of sires increases because more males contribute to the next generation (Shuster and Wade 2003; Shuster et al. 2013).

Detecting multiple paternity requires a sufficient number of polymorphic loci for both females and progenies. Once this condition is fulfilled, the test is easily performed in species having few progeny, for example, reptiles (Uller and Olsson 2008), birds (Birkhead and Moller 1995), sharks (Daly-Engel et al. 2006), and mammals (Say et al. 1999). In these cases, all progeny in a litter can be genotyped. However, in other cases, such as with endangered species having restrictions for sampling (e.g., turtles) or species having a large number of progeny in the same litter (i.e., invertebrates), only a small number or proportion can be analyzed. Female marine crabs carry thousands to millions of embryos before they release the free-swimming larvae (Hines 1991). Several crab species show multiple paternity; the first male can sire a high proportion of the brood, ranging from 50 to 98% of the progeny (Pardo et al. 2016). In these cases, only a small part of the progeny and the females in the population can be analyzed; thus, the technical problem arises of the minimum number of progeny and of females to be sampled in order to determine multiple paternity with high statistical power.

To detect the levels of multiple paternity, researchers commonly use the probability of detecting multiple mating (PrDM) proposed by Cobbs (1977) and Kichler et al. (1999) and coded by Neff and Pitcher (2002). This analysis allows estimating the statistical power to detect multiple paternity, including the following variables: (i) the number of microsatellites amplified, (ii) the proportion of each allele in the population, (iii) the number of offspring analyzed per litter, and (iv) the proportion of the progeny sired by the first male. Neff and Pitcher (2002) concluded that 10 embryos and four loci with four alleles in equal frequencies are necessary to have a statistical power of 95%.

Although the PrDM is a powerful tool to detect multiple paternity in a litter, this analysis neglects the estimation of the statistical power at the population level, where the experimental unit should be the clutch per female. This question is relevant for three main reasons: (i) not all females present multiple paternity in the population, (ii) when the study searches for temporal and geographical variation in multiple paternity, and (iii) to determine the likelihood of monogamy. For all these reasons, it is necessary to evaluate the statistical power at the population level.

In the case of the marine crab species, researchers used mainly few females (n = 10) and a large number of offspring per female (n = 100); however, the brood size (number of embryos per clutch) and the number of females in the population have not been considered. Considering this evidence, here, we developed an analytical method to assess the statistical power to detect multiple paternity considering brood size and the number of females sampled in a population, using crabs as a model.

Methods

Computation of statistical power We define statistical power as the probability of collecting at least one *sample* with larvae sired by several males (multiple paternity) among several samples each from a different brood. Hereafter, this probability will be denoted by P_{MP}. Also, male1 will refer to the male with the largest contribution of offspring in a brood, while male2 will refer to any contributing male other than male1 in cases of multiple paternity. Therefore, male2 may stand for several siring males. The parameters used are as follows:

NL	Number of larvae sampled in each brood
NB	Number of broods sampled
PBSPt	Proportion of broods with single paternity (sired by a single male)
PBMPt	Proportion of broods with multiple paternity (sired by several males)
TNLB	Estimated total number of larvae in a brood
PrLM1	Proportion of larvae sired by male1 in a brood sired by several males
PrLM2	Proportion of larvae sired by male2 in a brood sired by several males

Note that PrLM2 = 1 − PrLM1.

Procedure

a) P_{SPtAB}. First, we calculate the complementary probability which is the probability that not a single (0) *sample* was sired by more than one male. This is done in three steps.

i) P_{SPtMB} (single paternity for a *sample* from broods sired by several males). First, we estimate the probability that the NL larvae collected from a brood (sample) with multiple paternity were all sired by a single male. This calculation may be done analytically through exact combinatorics.

We define the following:
NM1: The total number of larvae from male1
NM2: The total number of larvae from male2

then
NM1 = TNLB × PrLM1
NM2 = TNLB × PrLM2

Single paternity in a sample collected from a brood sired by several males can happen in two ways:

Case 1: only offspring from male1 were collected.
Case 2: only offspring from male2 were collected.

The probability that we are seeking is the sum of the probabilities associated with each of these two possibilities. In other words, the probability that case 1 takes place is the ratio of all the choices of NM1 larvae (progeny from male1) among all NL larvae (all progenies). In standard combinatorics notation this following equation is written

$$P(\text{case } 1) = C_{NL}^{NM1} / C_{NL}^{(NM1+NM2)}$$

where C_k^n = the number of choices (combinations) of k elements among n elements
and $C_k^n = n!/[(n-k)\,!k!]$ ($n! = n \times (n-1) \times (n-2)\ldots \times 1$ reads n factorial)

A strictly analogous development for case 2 leads to:

$$P(\text{case } 2) = C_{NL}^{NM1} / C_{NL}^{(NM1+NM2)}$$

Summing over the two independent cases, we finally obtain

$$P_{SPtMB} = C_{NL}^{NM1} / C_{NL}^{(NM1+NM2)}$$

ii) P_{SPtB} (single paternity for a *sample* from broods sired by one or several males)

We then calculate the probability that any given sample was sired by a single male. Since the probability that all individuals of a sample from a brood sired by a single male were necessarily sired by a single male = 1, we have

$P_{SPtB} = PBMPt \times P_{SPtMB} + PBSPt \times 1$

iii) P_{SPtAB} (single paternity of samples for all broods)

This is estimated as $P_{SPtAB} = P_{SPtB}^{NB}$

b) Finally, we estimate the statistical power P_{MP} as $1 - P_{SPtAB}$.

If there are two sizes of samples, for example, 10 females with 20 progeny each and 20 females with 15 progeny each, the calculation of P_{SPtAB} generalizes to

$$P_{SPtAB} = P_{SPtB1}^{NB1} \times P_{SPtB2}^{NB2}$$

where

NL1 Size no. 1 of samples
NB1 Number of broods of size no. 1
NL2 Size no. 2 of samples
NB2 Number of broods of size no. 2
P_{SPtB1} Probability of single paternity for samples of size NL1 (from broods sired by one or several males)
P_{SPtB2} Probability of single paternity for samples of size NL2 (from broods sired by one or several males)

Clearly, this calculation can be generalized to any number of sample sizes as

$$P_{SPtAB} = P_{SPtB1}^{NB1} \times P_{SPtB2}^{NB2} \times P_{SPtB3}^{NB3} \times \ldots$$

This procedure was coded both in the programming language Maple 13 and in the R software (R Core Team 2015), the last posted as additional material in the journal webpage (Appendices 1 and 2).

It is important to note that the R software cannot evaluate our procedure with factorials larger than 50,000! This is problematic given that the estimated total number of progeny in a brood may well exceed 50,000. Therefore, we used and coded Ramanujan's approximation to ln(n!). This approximation greatly reduces the computation load while keeping very high levels of precision. We now describe this technique in some detail.

Ramanujan's approximation

$$ln(n!) \approx n\ln(n) - n + \frac{ln[n(1+4n(1+2n))]}{6} + \frac{\ln(\pi)}{2}$$

We denote Ramanujan's approximation of ln (n!) by RA(n)

The probability of detecting single paternity in a brood sired by several males may be re-written as follows

$$P_{SPMB} = \frac{e^{\ln\left(C_{NL}^{NM1}\right)} + e^{\ln\left(C_{NL}^{NM2}\right)}}{e^{\ln\left(C_{NL}^{NM1+NM2}\right)}}$$

or equivalently

$$P_{SPMB} = e^{\ln\left(C_{NL}^{NM1}\right) - \ln\left(C_{NL}^{NM1+NM2}\right)} + e^{\ln\left(C_{NL}^{NM2}\right) - \ln\left(C_{NL}^{NM1+NM2}\right)}$$

This last expression may then be approximated based on RA(n) and the following substitutions

$$C_k^n = n!/[(n-k)!k!]$$
$$\ln\left(C_k^n\right) = ln(n!) - ln[(n-k)!] - ln(k!)$$
$$\ln\left(C_k^n\right) = \mathrm{RA}(n) - \mathrm{RA}(n-k) - \mathrm{RA}(k)$$

Testing the statistical power: how many offspring, how many females? To test the statistical power, we used a number of microsatellites with large numbers of alleles, increasing the likelihood to detect multiple paternity in a brood. As stated by Neff and Pitcher (2002), four microsatellites with four alleles in equal frequency are enough to obtain good statistical power in a brood. To test different sample sizes (embryos and

females) from a population, we used the following prior: the researcher is able to genotype 1000 individuals from the progeny. We tested four different sampling schemes: (i) 10 females and 100 offspring per female, (ii) 20 females and 50 offspring per female, (iii) 50 females and 20 offspring per female, and (iv) 100 females and 10 offspring per female. We also considered three main features observed in crustaceans, all of which reduce the statistical power:

1. Marine crabs carry several thousand embryos in a clutch, a fact that reduces the statistical power to detect multiple paternity. For example, Hines (1991) reported Cancrid crab clutch sizes ranging from 780 embryos in *Cancer oregonensis* to 2,531,000 in *Cancer pagurus*. Considering these values, we used 1,000,000 embryos per female as a fixed value.
2. In natural populations, half of the females may carry progeny sired by at least two males (see Table 1). A reduction in the number of females carrying embryos sired by several males decreases statistical power. Therefore, our analysis considered the following four conditions: one half, one fourth, one eighth, and one sixteenth of females with multiple paternity.
3. In most cases, male2 (all males except male1) sires no more than 10% of the progeny in a brood (see Table 1). Considering that smaller proportions of contributions by male2 will generally reduce statistical power, we performed power calculations for the following proportions (male1/male2): 90:10; 95:5; 97.5:2.5; 98.75:1.25.

Results

The statistical power was estimated using the exact formulas coded in the Maple software and, also, using the Ramanujan's approximation in the R software. Both methods returned identical results, indicating that the Ramanujan's approximation allows accurate estimations of statistical power within the R software package.

The results of the analysis are summarized in Fig. 1. They show that all models reached values of statistical power greater than 99% when the first male sired 90% of the progeny and 50% of females had multiple paternity. However, when we used a low number of females and a large number of larvae, the statistical power was drastically reduced when both the number of larvae sired by male2 and the number of females having progeny sired by several males decreased. In the worst cases, when few females mate with multiple males and the first male sires 99% of the progeny, statistical power was low in all models tested. However, better values were reached when the number of females analyzed was increased in the model.

Discussion

This study developed a methodology that allows determining the statistical power of a sampling design in order to detect multiple paternity in species with high fertility, for which it is unfeasible to have a high representation of the offspring per female and a good representation of females from the whole population. The analysis pointed out the importance of using an adequate number of females to detect multiple paternity with high statistical power. We demonstrated that increasing the number of females analyzed increases the statistical power in the different combinations of number of females having progeny sired by several males and different proportions of embryos sired by male2.

This study points to the importance of the following considerations:

Table 1 Summary of paternity patterns for brachyuran species

Species	Number of females analyzed	Progeny sampled per female minimum-maximum	Paternity (prevalence)	Contribution of single male (%)	Reference
Chionoecetes opilio	25	50	Multiple (12%)	90 to 60	Sainte-Marie et al. (2008)
Dissodactylus primitivus	18	31–47	Multiple (60%)	92	Jossart et al. (2014)
Scopimera globosa	ED	ED	Multiple (ND)	94	Koga et al. (1993)
Ucides cordatus	10	8–16	Multiple (40%)	ND	Baggio et al. (2011)
Uca mjoebergi	9	80	Multiple (56%)	98	Reaney et al. (2012)
Metacarcinus magister	26	100	Multiple (40%)	98	Jensen and Bentzen (2012)
Cancer pagurus	18	40	Single	100	McKeown and Shaw (2008)
Metacarcinus edwardsii	54	10–20	Single	100	Pardo et al. (2016)

ED experimental design without genetic analysis, *ND* no data

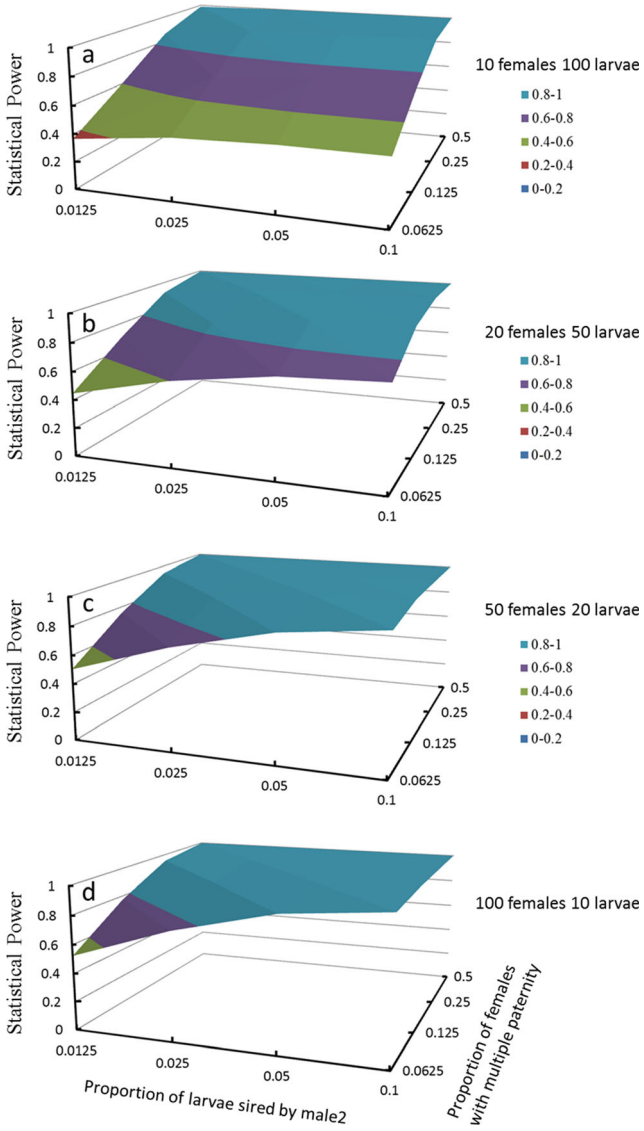

Fig. 1 Statistical power estimated with different proportions of female showing multiple paternity and different proportions of larvae sired by male2. **a** 10 females and 100 larvae. **b** 20 females and 50 larvae. **c** 50 females and 20 larvae. **d** 100 females and 10 larvae

1. In order to estimate power, one has first to estimate the mean number of progeny produced in a brood. This information is an important input since a large number of progeny per brood reduces the statistical power to detect multiple paternity.
2. From the analysis performed in this study, we propose 50 females and 20 embryos as an adequate sample size that guarantees high statistical power to detect multiple paternity even in species with 1,000,000 embryos per brood. However, we know that the sample sizes and number of females used in a study depend on both the sample accessibility (e.g., restrictions to collect a large number of ovigerous females) and the budget to amplify polymorphic genetic markers. Therefore, researchers will want to choose sampling designs that provide the desired level of statistical power at a minimal cost. Statistical power must then be assessed with a range of proportions of females showing multiple paternity and proportions of embryos sired by the first male.

As an example of the application of our method, we estimated the statistical power of a sampling scheme used with *Metacarcinus edwardsii* (Rojas-Hernandez et al. 2014; Pardo et al. 2016). It was estimated that this species may produce up to 1,000,000 embryos per clutch (Pardo pers. observations). The number of females was 54 and the number of embryos per clutch sampled was 10 (Pardo et al. 2016). Assuming 60% of females with single paternity and 98% of the progeny sired by the first male, the statistical power was estimated at 0.9835. As another example, we ran the procedure with *C. pagurus*. This species has a mean of 1,400,000 embryos per clutch (Hines 1991). Based on a sample of 18 females and 40 embryos per female (Mckeown and Shaw 2008), the power to detect multiple paternity was estimated at 0.9890. The estimates of the proportions of females with single paternity and the proportion of progeny sired by the first male were set as in the previous example. We conclude that, given our estimates of the proportion of females with multiple paternity and the proportion of embryos from male2, the sampling scheme of both studies ensured very high statistical power.

Finally, this analysis showed the importance of focusing on the experimental unit in the experimental design (clutch, nest, embryo per capsules, broods per female) in studies where multiple paternity is being tested. We also call attention to including information on the number of individuals in the progeny and the proportion of females expected to have multiple paternity in a population. Considering these factors and suitable genetic information, the mathematical routine presented here should provide a better estimation of the statistical power to detect multiple paternity in species with high fertility.

Acknowledgements DV thanks the financial support of Basal Grant PFB 023, ICM P05-002, and Millennium Nucleus NC120030. NRH thanks the IEB Master Thesis Fellowship. LMP thanks the financial support of the FONDECYT 1150388 and FONDAP-IDEAL 15150003. We thank Dr. Stephan Shuster and anonymous reviewers for constructive comments to the early version of this manuscript.

References

Baggio RA, Pil MW, Boeger WA, Patella LA (2011) Genetic evidence for multiple paternity in the mangrove land crab *Ucides cordatus* (Decapoda: Ocypodidae). Mar Biol Res 7:520–524

Birkhead TR, Moller AP (1995) Extra-pair copulation and extra-pair paternity in birds. Anim Behav 49:843–848

Cobbs G (1977) Multiple insemination and male sexual selection in natural populations of *Drosophila pseudoobscura*. Am Nat 111:641–656

Daly-Engel TS, Grubbs RD, Holland KN, Toonen RJ, Bowen BW (2006) Assessment if multiple paternity in single litters from three species of carcharhinid sharks in Hawaii. Environ Biol Fish 76:419–424

Griffith SC, Owens IPF, Thuman KA (2002) Extra pair paternity in birds: a review of interspecific variation and adaptive function. Mol Ecol 11:2195–2212

Hines AH (1991) Fecundity and reproductive output in nine species of Cancer crabs (Crustacea: Brachyura: Cancridae). Can J Fish Aquat Sci 48:267–275

Jensen PC, Bentzen P (2012) A molecular dissection of the mating system of the Dungeness crab, *Metacarcinus magister* (Brachyura: Cancridae). J Crust Biol 32:443–456

Jossart Q, Wattier R, Kastally C, Aron S, David B, De Ridder C, Rigaud T (2014) Genetic evidence confirms polygamous mating system in a crustacean parasite with multiple hosts. PLoS One 9:e90680

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

Koga T, Henmi Y, Murai M (1993) Sperm competition and the assurance of underground copulation in the sand-bubbler crab *Scopimera globosa* (Brachyura: Ocypodidae). J Crust Biol 13:134–137

McKeown NJ, Shaw PW (2008) Single paternity within broods of the brown crab *Cancer pagurus*: a highly fecund species with long-term sperm storage. Mar Ecol Progr Ser 368:209–215

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

Pardo LM, Riveros MP, Fuentes JP, Rojas-Hernandez N, Veliz D (2016) An effective sperm competition avoidance strategy in crabs drives genetic monogamy despite evidence of polyandry. Behav Ecol Sociobiol 70:73–81

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Reaney L, Maurer G, Backwell PY, Linde C (2012) Paternity analysis of two male mating tactics in the fiddler crab, *Uca mjoebergi*. Behav Ecol Sociobiol 66:1017–1024

Rojas-Hernández N, Veliz D, Pardo LM (2014) Use of novel microsatellite markers for population and paternity analysis in the commercially important crab *Metacarcinus edwardsii* (Brachyura: Cancridae). Mar Biol Res 10:839–844

Sainte-Marie B, Gosselin T, Sévigny J-M, Urbani N (2008) The snow crab mating system: opportunity for natural and unnatural selection in a changing environment. Bull Mar Sci 83:131–161

Say L, Pontier D, Natoli E (1999) High variation in multiple paternity of domestic cats (*Felis catus* L.) in relation to environmental conditions. Proc Roy Soc Lond B 266:2071–2074

Shuster SM, Wade MJ (2003) Mating systems and strategies. Princeton University Press, Pinceton, NJ, 522 pp

Shuter SM, Briggs WR, Dennis PA (2013) How multiple mating by females affects sexual selection. Philos Trans R Soc Lond Ser B Biol Sci 368:20120046

Uller T, Olsson M (2008) Multiple paternity in reptiles: patterns and processes. Mol Ecol 17:2566–2580