Journal of Heredity 2009:100(2):197–205
doi:10.1093/jhered/esn095
Advance Access publication November 4, 2008

Assessing Parent Numbers from Offspring Genotypes: The Importance of Marker Polymorphism

KRISTINA M. SEFC AND STEPHAN KOBLMÜLLER

From the Department of Zoology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria (Sefc); the Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden (Koblmüller).

Address correspondence to K. M. Sefc at the address above, or e-mail: kristina.sefc@uni-graz.at.

Abstract

Methods to infer parent numbers from offspring genotypes either determine the minimum number of parents required to explain alleles and multilocus genotypes detected in the offspring or use models to incorporate information on population allele frequencies and allele segregation. Disparate results by different approaches suggest that one or perhaps all methods are subject to bias. Here, we investigate the performance of minimum parent number estimates, maximum likelihood, and Bayesian analyses (programs COLONY and PARENTAGE) with respect to marker information content in simulated data sets without knowledge of parental genotypes. Offspring families of different sizes were assumed to share one parent and to be sired by 1 or 5 additional parents. All methods committed large errors in terms of underestimation (minimum value) and overestimation (COLONY), or both (PARENTAGE) of parent numbers, unless the data were highly informative, and their relative performances depended on full-sib group sizes and sire numbers. Increasing the number of markers with low gene diversity ($H_e \leq 0.68$) yielded only slow improvement of the results, but all 3 methods performed well with 5–7 markers of $H_e = 0.84$. We emphasize the importance of high marker polymorphism for inferring parent numbers and individual parent contributions, as well as for the detection of monogamous reproduction.

Key words: *GERUD, monogamy, multiple mating, parentage, paternity*

Studies of animal and plant mating systems, sperm and pollen competition, parental investment in relation to genetic parentage, pre- and postzygotic selection, the distribution of reproductive success, and several other issues of evolutionary and ethological interest depend on the determination of the number and identity of reproducing individuals (e.g., Bernasconi 2003; Uller and Olsson 2008). Molecular methods contributed significantly to the field and increased our knowledge far beyond the amount of information that can, for example, be obtained from behavioral observation of breeding animals and plant pollinators (e.g., Avise et al. 2002; Griffiths et al. 2002; Llaurens et al. 2008). Accordingly, methods for the analysis of genetic data with respect to relatedness and parentage inference have been developed and refined in recent years (e.g., Marshall et al. 1998; Neff et al. 2002; Wang 2004; Jones 2005; Jones et al. 2007). One of the many applications of genetic and allozyme data to parentage studies is the reconstruction of the number of parents from the genotypes of animal offspring or plant seeds and seedlings. Assuming that all offspring in a group have one of their parents in common, which is the case for seeds collected from inflorescences and fruits as well as for broods of animals in which one sex mates with several individuals of the other sex, the data analysis consists in first deducing the genotype of the shared parent (if it has not been sampled) and then determining the number of additional parents contributing alleles to the offspring group. The power to exclude unrelated individuals from parentage and to discriminate different parents increases with increasing levels of polymorphism of the genetic markers, as it becomes more likely that different parents transmit different alleles to their respective offspring (Neff and Pitcher 2002). Such highly polymorphic markers are not always readily available, and likelihood methods have been developed to account for the possibility of allele sharing between different parents using population allele frequencies and Mendelian expectations for the ratios of alleles in full-sib groups. With these algorithms, 2 offspring sharing frequent alleles are not necessarily inferred to also share the same parent (e.g., Meagher 1986; Emery et al. 2001; Smith

et al. 2001; Wang 2004; Hain and Neff 2007). Hence, the minimum number of parents required to explain the genotypes detected in the offspring is often smaller than the maximum likelihood estimate of parent numbers (Campbell 1998; Mäkinen et al. 2007; Portnoy et al. 2007; Sefc et al. 2008).

Highly disparate results obtained from the estimation of parent numbers by parsimonious minimum sire number reconstructions and maximum likelihood analyses of empirical data (Sefc et al. 2008; Hermann C, Koblmüller S, Sefc KM, unpublished data) and furthermore the failure to retrieve the correct answers from simulated data with both methods, when the markers were not sufficiently informative (Sefc et al. 2008), prompted us to examine the performance of different methods for paternity reconstruction in more detail. In the present study, we investigate the effect of marker information content (number of markers and their diversity) on reconstructions of minimum parent number, maximum likelihood, and Bayesian estimates (programs COLONY and PARENTAGE) in simulated data sets without knowledge of parental genotypes. We refer to the simulated data as the genotypes of broods, in which offspring share a common mother and were sired by one or several males, but the same type of data is encountered when broods are sired by a single father and several mothers or when plant inflorescences or fruits are tested for pollination by multiple donors.

Materials and Methods

Parentage was reconstructed from 13 data sets consisting of different numbers of loci (n = 3, 5, 7, 9, and 11 loci) and with 3 different levels of polymorphism per marker (gene diversity H_e = 0.57, 0.68, and 0.84). The exclusion probabilities achieved with each marker set (P_1, when one parent is known, and P_0, when neither parent is known) were calculated in GERUD vs 2.0 (Jones 2005) and are given in Table 1. The number of alleles and allele frequencies, which were used for the simulation of brood genotypes and to represent the population from which broods were sampled, were as follows: for H_e = 0.57, 5 alleles at frequencies of 0.54098, 0.36067, 0.06557, 0.02732, and 0.00546; for H_e = 0.68, 8 alleles at frequencies of 0.41919, 0.32576, 0.14393, 0.09091, 0.0101, 0.00505, 0.00253, and 0.00253; and for H_e = 0.84, 12 alleles at frequencies of 0.26053, 0.17105, 0.15789, 0.12368, 0.10526, 0.09737, 0.02895, 0.01842, 0.01316, 0.01053, 0.01053, and 0.00263; these are values that were observed at 3 microsatellite loci in a large natural population sample.

The simulation of broods assumed that all offspring shared one parent and that 5 different individuals acted as the second parent, for example, such as when a female mated with 5 males, each of whom sired a proportion of the clutch. Two brood sizes, n = 80 and n = 25, were simulated. Paternity contributions were either evenly distributed among the 5 sires, with each male siring 16 or 5 offspring in the large and small broods, respectively, or had a skewed distribution with the primary sire accounting for 40 (large broods) or 13 (small broods) of the offspring and the remaining 4 males each siring 10 (large broods) or 3 (small broods) young. "Large" and "small" broods therefore contain full-sib groups of different sizes, whereby larger full-sib groups provide more information to reconstruct the genotypes of their parents. Furthermore, monogamous broods (1 mother and 1 father) with 25 and 80 offspring were simulated.

To generate brood genotypes, parental genotypes were first assembled according to the population allele frequencies. For example, for data consisting of 5 markers, each with H_e = 0.68, 10 alleles (2 per locus) were randomly drawn for each assumed parent from the above allele distribution of the marker with H_e of 0.68. An offspring's genotype was then created by randomly drawing 1 allele per locus from each of its parents. For each of the investigated combinations of brood type and data set composition, 100 replicate broods were generated and analyzed.

Because all offspring in a brood shared their mother, the reconstructed number of full-sib groups in a brood was taken as estimate of the number of different sires contributing to the brood. For each of the multiply sired broods, we calculated maximum likelihood estimates of sire numbers (COLONY estimates), Bayesian estimates (PARENTAGE estimates), and reconstructions of the minimum number of sires required to explain offspring genotypes (MIN estimates). For monogamous broods, only COLONY and PARENTAGE estimates were calculated because the minimum sire number estimate could not be lower than 1 anyway. As neither mutations nor typing errors were included in the genotype simulations, sibship reconstruction

Table 1. Exclusion probabilities achieved by the marker sets used in brood simulations when one parent is known (P_1) and when neither parent is known (P_0)

# Loci H_e	3 Loci	5 Loci	7 Loci	9 Loci	11 Loci
0.57	P_1 = 0.6404	P_1 = 0.8181	P_1 = 0.9080	P_1 = 0.9535	P_1 = 0.9669
	P_0 = 0.4192	P_0 = 0.5957	P_0 = 0.7185	P_0 = 0.8041	P_0 = 0.8365
0.68	P_1 = 0.8208	P_1 = 0.9431	P_1 = 0.9819	P_1 = 0.9942	P_1 = 0.9968
	P_0 = 0.6101	P_0 = 0.7919	P_0 = 0.8889	P_0 = 0.9407	P_0 = 0.9567
0.84	P_1 = 0.9686	P_1 = 0.9969	P_1 = 0.9997	n/a	n/a
	P_0 = 0.8870	P_0 = 0.9736	P_0 = 0.9938		

H_e, expected heterozygosity per locus; # loci, number of loci, each with the indicated H_e, making up the data set. n/a, not applicable.

in the program COLONY vs 1.0 (Wang 2004) was carried out with error rates set to zero. For maximum likelihood estimates of the number of sires contributing to a brood, COLONY was run as described in the manual, using the population allele frequencies given above. To obtain MIN estimates, the population allele frequencies were set to 0.001 for all alleles present in the offspring, and an additional allele (not present in the offspring) was given a high frequency in order to make allele frequencies add up to 1. When alleles found in offspring are assigned low population frequencies, it is more likely that a shared allele represents shared paternity, and the maximum likelihood estimate returns the minimally required number of sire genotypes (Sefc et al. 2008). To confirm convergence of the MIN method on the minimum sire number, 7 monogamous data sets were analyzed. The necessity to employ this modification of COLONY, which has the disadvantage that it may occasionally overestimate the minimum value (see Results), was given by the need to automate the large number of analyses. When analyses can be run individually and no more than 6 sires contributed to the brood, minimum sire numbers are better calculated by the program GERUD (Jones 2005), which tests increasing numbers of sires for compliance with the offspring genotypes and is therefore guaranteed to arrive at the true minimum.

The Bayesian parentage reconstruction implemented in PARENTAGE (Emery et al. 2001) evaluates probability distributions of parent numbers and their relative contributions conditional on observed offspring genotypes and population allele frequencies and prior assumptions. Because of long computation times (exceeding 2 h per run with the more informative data sets), not all the generated data sets, and only 50 replicates of each, were analyzed with PARENTAGE. The analyzed data sets were as follows: both large and small multiply sired broods with skewed paternity distributions, as well as large monogamous broods, with data for 3, 5, and 7 loci at H_e of 0.57, 0.68, and 0.84 and small monogamous broods with data for 3 and 5 loci at H_e of 0.57, 0.68, and 0.84. To improve the mixing properties of the Markov Chain, a low probability of more than 1 mother was set by using a prior distribution with a mean of 1 and a standard deviation of 0.1 (Emery et al. 2001; PARENTAGE manual). The prior range of possible fathers was set to 1–15. In previous studies, results were shown to be robust to changes in priors for father number and paternity share (Bretman and Tregenza 2005; Beveridge et al. 2006). The simulated data contain no genotyping errors and mutations, but because the program does not allow a mutation rate of zero, the prior for the mutation rate was set to the low value of 10^{-10}. The here chosen Markov Chain settings of 5000 iterations with a burn-in of 5000 iterations and a thinning interval of 400 are commonly employed by users of the program (e.g., Bretman and Tregenza 2005; Beveridge et al. 2006; Simmons et al. 2007; Frentiu and Chenoweth 2008).

Scripts to generate brood genotypes and COLONY and PARENTAGE input files, run the programs, and parse the output files were written in PERL programming language.

Results

Errors in sire number estimates by the MIN and COLONY analyses were in most cases due to under- and overestimation, respectively, of the true value, whereas PARENTAGE estimates deviated from the true value in both directions (Figures 1 and 3; Supplementary Figure). The allocation of the correct number of offspring to each contributing sire is even more difficult than the estimation of the number of sires involved in a brood. The percentage of analyses, in which the proportions of paternal contributions to the broods were correctly reconstructed, was considerably lower than the percentage of analyses with accurate estimates of sire number, except with the most powerful marker set (Figure 1; Supplementary Figure).

In a small number of analyses (0.8%), the MIN estimator overestimated the true value by 1 (and once by 2) sire. Apparently, the frequency of 0.001 assigned to each observed allele, albeit very low, did not force the likelihood algorithm to minimize the sire number estimate in all cases. Hence, even when our MIN estimate was equal to or smaller than the true sire number, the minimally necessary number of sires to explain all offspring genotypes may be still smaller, such that our analyses may underrate the degree to which the "accurate" minimum numbers would underestimate the true number of sires. However, 2 lines of evidence indicate that the MIN analyses in COLONY will in most cases arrive at the accurate minimum sire numbers. In a previous study, the MIN estimates obtained from the modified COLONY input were identical to minimum sire number reconstructions in GERUD (Sefc et al. 2008). Moreover, in the present study, several data sets of monogamous broods were analyzed with the MIN method in COLONY in order to confirm that the method indeed returned the minimally possible sire number (known to be 1 in monogamous broods). Seven data sets, in which the maximum likelihood estimate of COLONY overestimated sire number, were chosen for this check, and the MIN method correctly estimated a single sire in all replicates (large broods with data from 3 and 5 loci and H_e of 0.57 and 0.68, and 3 loci with H_e of 0.84). Although we do not advocate our modification of the COLONY program to obtain MIN estimates in cases, where it is possible to use the program GERUD, one potential advantage of the MIN sire number reconstruction in COLONY compared with GERUD lies in the ability of the COLONY model to account for genotyping errors and mutations in empirical data sets, which could cause spurious assumptions of additional sires by GERUD analyses.

Paternity Reconstruction from Large Full-Sib Groups ("Large Broods")

In large broods, the distribution of paternity among sires had little effect on the outcome of paternity reconstruction with the MIN and COLONY methods, and success rates were similar in broods with a primary sire and in broods with equally contributing sires (Figure 1A–C and Supplementary

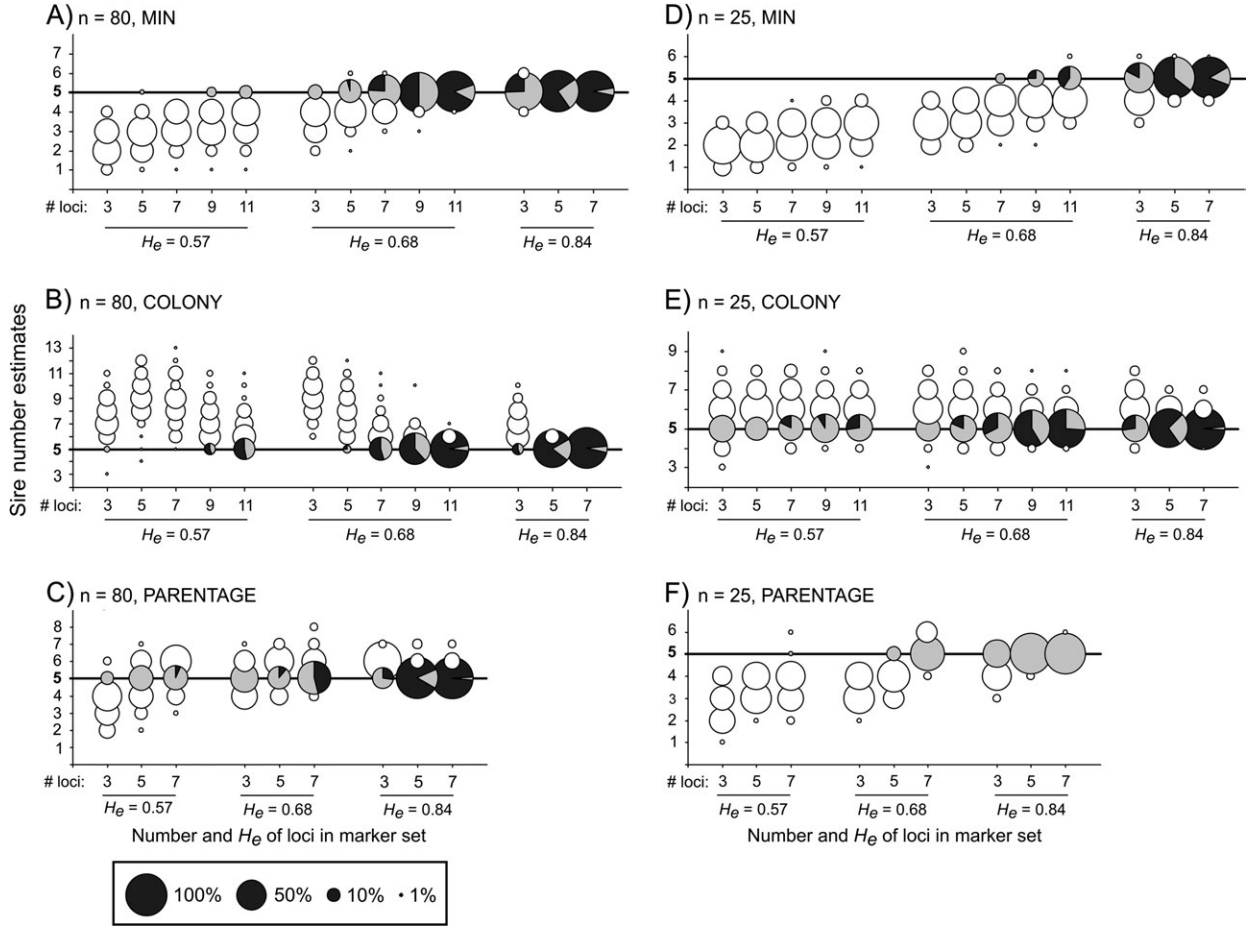

Figure 1. Sire number estimation from simulated brood genotypes with skewed distribution of paternal contributions. Bubble areas represent the percentage of replicate broods, for which minimum (MIN), maximum likelihood (COLONY), and Bayesian (PARENTAGE) analyses returned the sire numbers given on the *y* axis. The horizontal line marks the correct estimate of 5 sires, and bubbles representing the proportions of correct estimates are shaded gray and black. Gray areas indicate the portion of analyses yielding correct sire numbers but erring in the number of offspring contributed by each sire; black areas correspond to the proportion of analyses, in which both sire number and contributions were correctly inferred. Marker sets consisted of 3–11 loci with gene diversities (H_e) of 0.57 and 0.68 and 3–7 loci with $H_e = 0.84$. (**A**–**C**) MIN, COLONY, and PARENTAGE analyses of 80 offspring, where the primary father sired 40 young and each of the additional 4 fathers sired 10 young. (**D**–**F**) MIN, COLONY, and PARENTAGE analyses of 25 offspring, where the primary father sired 13 young and each of the additional 4 fathers sired 3 young.

Figure). Apparently, there was little difference in the amount of information available for paternity reconstruction between 16 offspring per sire in the even broods and 10 offspring per secondary sire in the skewed broods. PARENTAGE analyses were conducted on skewed broods only.

Markers with low polymorphism ($H_e = 0.57$) performed poorly in paternity reconstruction with all 3 methods even when a high number ($n = 7$–11) of such markers was used (Figures 1A–C and 2A). Unexpectedly, the rate of correct paternity inference by COLONY decreased, when the number of markers was increased from 3 to 5 and 7. Only sets of 9 and 11 markers yielded better COLONY results than a set of 3 markers, but the proportion of analyses with correctly inferred paternity remained low, and deviations of sire number estimates from true values remained high (Figures 1B and 2A). The proportion of correct sire number estimates was higher in the PARENTAGE program but still remained below 50% with up to 7 markers (Figure 1C). Due to the long computation times, data sets with more than 7 loci were not analyzed with PARENTAGE. The average deviations between PARENTAGE estimates and true sire numbers were lower than with the other 2 methods (Figure 2A).

With moderately polymorphic markers ($H_e = 0.68$), at least 9 markers were required for the MIN method and 11 for the COLONY method to identify the true sire number in >80% of replicates and minimize the deviation of the

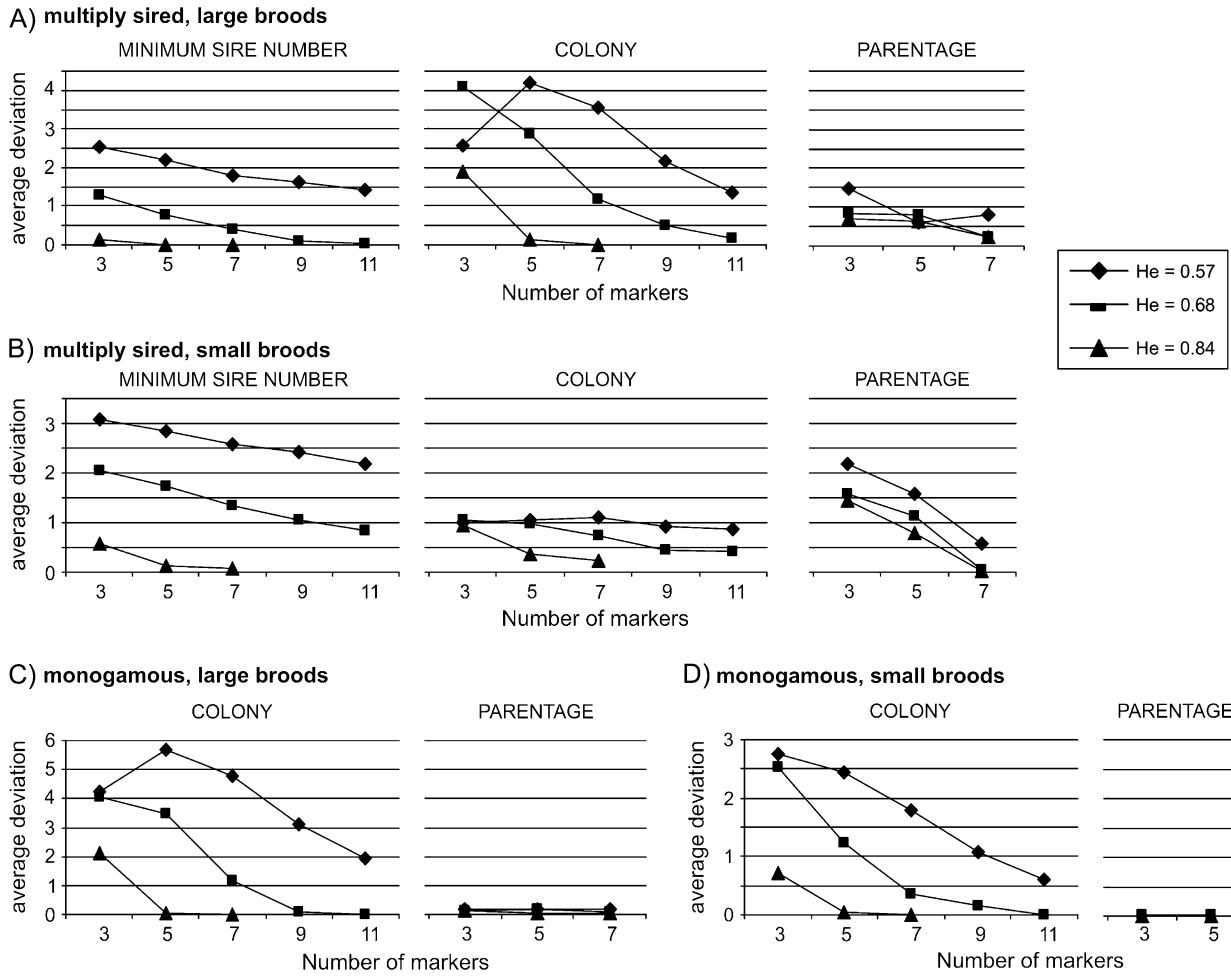

Figure 2. Average deviations between true sire numbers and numbers estimated from simulated brood genotypes. Deviations were averaged across the 100 replicates analyzed for each brood type and marker set with the MIN and the COLONY methods and the 50 replicates analyzed with PARENTAGE. (**A**) Broods of 80 offspring sired by a primary father (40 offspring) and 4 additional sires (10 offspring each). (**B**) Broods of 25 offspring sired by a primary father (13 offspring) and 4 additional sires (3 offspring each). (**E**) Monogamous broods with 80 offspring. (**F**) Monogamous broods with 25 offspring.

incorrect estimates (Figures 1A–B and 2A). The PARENTAGE estimates based on 3, 5, and 7 loci were correct in 40, 28, and 52% of the replicates, respectively, with only modest improvement of estimates with increasing marker numbers (Figures 1C and 2A). Although not optimal, the PARENTAGE method performed better with small data sets (3–7 markers) than the other 2 methods in terms of the proportion of correct estimates of sire number and paternal contribution and of the average deviation from the true sire number. Interestingly, fewer markers with H_e of 0.57 and 0.68 caused PARENTAGE to underestimate sire numbers, whereas overestimation was more frequent with data sets containing more markers.

With highly polymorphic markers (H_e = 0.84), a set of 3 markers was sufficiently informative to infer sire number in >80% of the replicates by the MIN method, whereas 5–7 markers were needed to achieve the same success rate with the COLONY and PARENTAGE methods (Figure 1A–C). With the most informative marker set (7 loci with H_e = 0.84), both sire number and individual paternal contribution were estimated correctly in at least 95% of the replicates by MIN and COLONY analyses and in >80% of the replicates by PARENTAGE.

Except in the analyses with the most powerful marker sets, the COLONY method overestimated sire numbers by up to 100%, whereas the deviations of sire number estimates from the true value were lower with the MIN and PARENTAGE methods (Figures 1A–C and 2A). When interested in mating frequencies (e.g., Song et al. 2007), it may in fact be more important to minimize the total deviation across analyses than to maximize the rate at which the true values are recovered at the cost of large errors for some broods. In this case, the MIN method and the PARENTAGE model may be appropriate choices for analyses of large broods and moderately polymorphic markers.

Paternity Reconstruction from Small Full-Sib Groups ("Small Broods")

Given that the amount of information available for paternity reconstruction increases with increasing numbers of offspring per sire, one would expect to obtain better results from the large broods, where each of the males sired at least 10 offspring, than in the small broods, where some sires are represented by only 3 offspring. Indeed, the MIN method produced fewer correct inferences of sire numbers and contributions and greater deviations of sire number estimates from the true value in the small broods than in the large broods (Figures 1A,D and 2A,B). MIN analyses with 3–11 markers of $H_e = 0.57$ completely failed to reconstruct the correct sire number, and accurate inference of sire number in >80% of the replicates required 5–7 markers of $H_e = 0.84$. Similar results were obtained with the PARENTAGE program, which produced fewer correct estimates from small broods except when highly informative data sets were used (Figure 1F). Although the rates of correct sire number estimates by PARENTAGE were comparable to those achieved by the MIN method, none of the PARENTAGE runs returned correct estimates of individual paternal contributions (Figure 1F).

The COLONY method generally performed better in the small broods than it did in the large broods in terms of the deviations between sire number estimates and true values (Figure 2B), although not always in terms of the proportion of correct sire number inferences (Figure 1B,E). In particular, accurate COLONY sire number estimates based on the less informative marker sets were obtained more frequently in the small than in the large broods, whereas the rates of accurate estimates based on highly informative marker sets were generally lower in the small than in the large broods, the latter conforming to expectations on the effect of brood sample size (Wang 2007). Moreover, the proportions of correct estimates based on the less informative marker sets were considerably higher, and deviations from true values were lower, by the COLONY than by the PARENTAGE and the MIN methods (Figures 1D–F and 2B). Nonetheless, only the most informative marker set recovered the correct sire number in >80% of the replicates (Figure 1E).

Both the MIN and the COLONY methods performed somewhat worse in small broods with skewed sire contributions, in which secondary sires were represented by only 3 offspring each (Figure 1D–E), than in those broods, to which each of the sires contributed 5 offspring (Supplementary Figure). A negative effect of reproductive skew on paternity reconstruction was also observed in other simulation studies (Neff et al. 2002; Myers and Zamudio 2004). Only skewed broods were analyzed with PARENTAGE.

Monogamous Broods

An accurate reconstruction of the minimum number of sires required to explain genotypes of "monogamous broods" (i.e., broods sired by a single male and a single female) will always arrive at the correct sire number of $n = 1$, but unless the exclusion probability of the marker set is very high, this does not prove monogamy. In fact, in the simulated broods with 5 sires, a small proportion of the MIN and PARENTAGE results were consistent with monogamy (Figure 1A,D,F). The risk of falsely inferring monogamy from multiply sired broods by the MIN method is closely related to the exclusion probability of the data and can be assessed, for example, in the program GERUDSIM (Jones 2005), by simulating and analyzing user-defined broods and marker sets.

Vice versa, monogamous broods may be erroneously assigned multiple parents by maximum likelihood and Bayesian models. Indeed, as in the above COLONY analyses of multiply sired broods, sire number was greatly overestimated in monogamous broods unless highly informative marker sets were used, especially when the broods were large (Figures 3A,C and 2C,D). At least 9 markers of $H_e = 0.68$, or 5 markers of $H_e = 0.84$, were required for COLONY to detect monogamy in >80% of the replicates in both large and small broods.

In contrast, monogamy of broods was correctly inferred by PARENTAGE in >80% of replicates with all tested data sets (Figure 3B,D). Although this result is highly desirable, it may be linked to the propensity of PARENTAGE to underestimate sire numbers from small data sets (see above), which makes results converge on a single sire, rather than represent superior capability of sire number reconstruction.

Discussion

Most importantly, our analyses suggest that paternity inference from brood genotypes requires highly informative marker sets and that low levels of marker polymorphism cannot be easily compensated for by increasing the number of such markers. Moreover, even sophisticated methods including allele segregation and population allele frequency information into the estimation of sire number were unable to arrive at correct solutions when the markers were not sufficiently informative. Hence, it will in many circumstances prove more efficient to invest in the development of highly polymorphic markers than to sufficiently increase the number of markers with little or moderate diversity. These findings may be of particular importance for parentage studies in plants, where allozyme markers have long remained the predominant tool (Bernasconi 2003) and have only recently been superseded by the employment of highly polymorphic microsatellite markers (Reusch 2000; Teixera and Bernasconi 2007; Llaurens et al. 2008). However, even with microsatellites, genetic diversity values above 0.8 (required to obtain correct results in our analyses of 5 and more loci) may, at least in some species or populations, not be encountered at many loci. In fact, only few microsatellite-based paternity studies employ either highly polymorphic markers (e.g., Adams et al. 2005; Mackiewicz et al. 2005; Portnoy et al. 2007; Teixera and Bernasconi 2007; Wilson and Martin-Smith 2007) or more than 7 loci (e.g., Reusch 2000; Myers and Zamudio 2004; Beveridge et al. 2006;

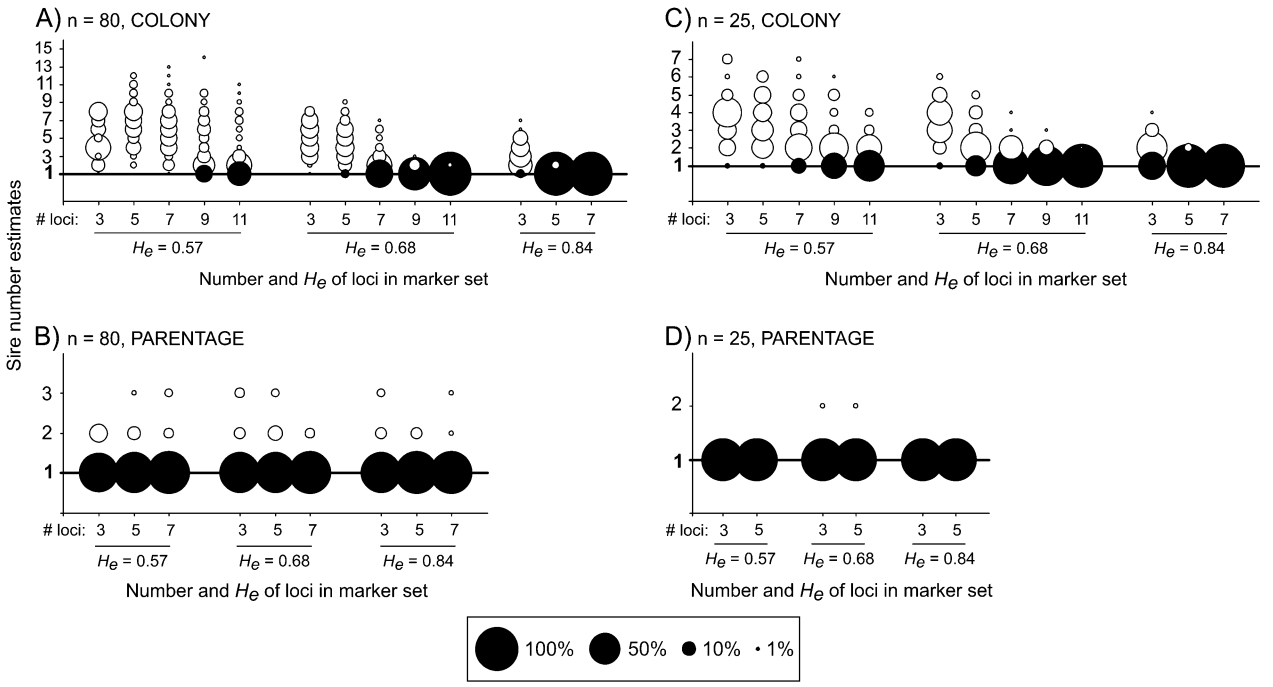

Figure 3. Sire number estimation from monogamous broods. Bubble areas represent the percentage of replicate broods, for which maximum likelihood (COLONY) and Bayesian (PARENTAGE) analyses returned the sire numbers given on the *y* axis. Black bubbles represent correct estimates, that is, a single sire (also marked by the horizontal line). (**A**) COLONY and (**B**) PARENTAGE analyses of 80 offspring. (**C**) COLONY and (**D**) PARENTAGE analyses of 25 offspring.

Herbinger et al. 2006), whereas the majority of studies are being conducted with only 3–5 markers of oftentimes moderate diversity. Investigators should bear in mind that, without more informative data, estimates of parent numbers and relative paternal contributions may be approximate (Sefc et al. 2008).

It is also noteworthy that marker sets with higher exclusion probabilities did not necessarily produce better sire number estimates, depending on analysis method and family structure. For example, although exclusion probabilities increased rapidly between 3 and 7 markers of $H_e = 0.57$ (Table 1), COLONY inferences of sire numbers deteriorated (Figure 2). Furthermore, a set of 3 markers with $H_e = 0.84$ had lower exclusion probabilities than 7 markers with $H_e = 0.64$ (Table 1) but yielded more accurate MIN estimates, particularly for small broods (Figures 1 and 2).

The positive correlation between brood size and corresponding COLONY estimates of sire number is an artifact of the analysis method but could be mistaken, in a methodological context, for a failure to detect all contributing sires in small brood samples (Hain and Neff 2007), or, in a biological context, for an indication of increasing offspring numbers with increasing numbers of mates (Bateman 1948; Neff et al. 2008). Generally, both the overestimation of sire numbers and the failure to detect the full number of extra sires may compromise conclusions on the extent of polygamous mating, sneaking rates, and other alternative reproductive behaviors (e.g., Chapman et al. 2004). In particular, when methods biased toward sire number underestimation (e.g., the MIN and PARENTAGE methods) are applied to moderately informative data sets, inferences of monogamy could also be obtained from broods with more than a single sire.

Our results prompt no general recommendation of one of the methods over the others because their performances, relative to each other, depended on brood size, the true number of sires, and the distribution of paternal contributions between sires. Overall, the 3 methods fared poorly with fewer, and less polymorphic, markers and estimated sire numbers similarly well when the markers were highly informative. The COLONY method performed best in reconstructing the individual contributions of the inferred sires (Figure 1; Supplementary Figure). Importantly, useful information can be drawn from comparisons between estimates of the different methods, as disparate results indicate that there may be insufficient information for accurate paternity reconstruction. In these cases, the true sire number will probably lie within the range of the different estimates (see also Sefc et al. 2008). In contrast, congruency between estimates from methods biased in opposite directions provides strong support for the obtained estimate.

Another maximum likelihood parentage reconstruction method, implemented in the program PEDIGREE (Herbinger 2005), yielded similar results as COLONY from empirical data (Herbinger et al. 2006; Sefc et al. 2008) and displayed similar biases in paternity reconstruction from

simulated data (Sefc et al. 2008). The tendency of the PEDIGREE algorithm to split large full-sib groups into subgroups, which causes a positive correlation between brood size and inferred sire number, can be compensated by analysis of parameter settings penalizing group splitting (Herbinger et al. 2006), such that the minimum number of sires can be approached with higher penalties, which then entails the risk of underestimating the true value. The settings required to arrive at the true number of sires vary with the composition of the broods (Sefc et al. 2008), which makes the choice of the settings a critical step of the PEDIGREE analyses.

Jones et al. (2007) compared parentage inference in COLONY, PARENTAGE, and NEST (their own Bayesian model, which assigns reproductive success to different categories of candidate parent individuals based on offspring and candidate parent genotypes), and reconstructions of the minimum numbers of parents per nest, in an investigation of the reproductive success of different categories of candidate parents. Their data comprised 23 nests, each represented by 48 offspring genotyped at 5 loci with H_e values between 0.57 and 0.88 (mean $H_e = 0.74$, $P_0 = 0.92$, $P_1 = 0.99$; Fiumera et al. 2002). On average, COLONY estimated twice as many mothers per nest as the other 3 methods. Although the true number of parents per nest was unknown, it was concluded that COLONY overestimated parent numbers. This agrees with our findings for this level of marker polymorphism. Likewise, the congruency between the minimum parent number and the PARENTAGE reconstructions (Jones et al. 2007) is consistent with our observation that MIN and PARENTAGE methods perform similarly well and display a comparable bias in small broods. Based on NEST analyses of simulated data and weighing the effort and benefits associated with increased sampling of loci, different nests, or offspring per nest, Jones et al. (2007) propose to analyze no more than 3 or 4 loci in order to offset the costs associated with the genotyping of many broods. Although this number of markers may indeed be sufficient to reconstruct parentage when candidate parents are included in the data set, our analysis indicates that sire number reconstruction without parental information requires more, or more polymorphic, loci. Furthermore, although none of the methods tested in our study was able to reconstruct sire numbers from moderately and little informative data sets, the employment of different methods appears to be a useful way to assess the reliability of the obtained results.

Supplementary Material

Supplementary materials can be found at http://www.jhered.oxfordjournals.org/.

Funding

Austrian Research Fund (P17380) to K.M.S.

References

Adams EM, Jones AG, Arnold SJ. 2005. Multiple paternity in a natural population of a salamander with long-term sperm storage. Mol Ecol. 14:1803–1810.

Avise JC, Jones AG, Walker D, DeWoody JA. 2002. Genetic mating systems and reproductive natural histories of fishes: lessons for ecology and evolution. Annu Rev Genet. 36:19–45.

Bateman AJ. 1948. Intra-sexual selection in *Drosophila*. Heredity. 2:349–368.

Bernasconi G. 2003. Seed paternity in flowering plants: an evolutionary perspective. Perspect Plant Ecol Evol Syst. 6:149–158.

Beveridge M, Simmons LW, Alcock J. 2006. Genetic breeding system and investment patterns within nests of Dawson's burrowing bee (*Amegilla dawsoni*) (Hymenoptera: Anthophorini). Mol Ecol. 15:3459–3467.

Bretman A, Tregenza T. 2005. Measuring polyandry in wild populations: a case study using promiscuous crickets. Mol Ecol. 14:2169–2179.

Campbell DR. 1998. Multiple paternity in fruits of *Ipomopsis aggregata* (Polemoniaceae). Am J Bot. 85:1022–1027.

Chapman DD, Prödohl PA, Gelsleichter J, Manire CA, Shivji MS. 2004. Predominance of genetic monogamy of females in a hammerhead shark, *Sphyrna tiburo*: implications for shark conservation. Mol Ecol. 13:1965–1974.

Emery AM, Wilson IJ, Craig S, Boyle PR, Noble LR. 2001. Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. Mol Ecol. 10:1265–1278.

Fiumera AC, Porter BA, Grossman GD, Avise JC. 2002. Intensive genetic assessment of the mating system and reproductive success in a semi-closed population of the mottled sculpin, *Cottus bairdi*. Mol Ecol. 11: 2367–2377.

Frentiu FD, Chenoweth SF. 2008. Polyandry and paternity skew in natural and experimental populations of *Drosophila serrata*. Mol Ecol. 17: 1589–1596.

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

Hain TJA, Neff BD. 2007. Multiple paternity and kin recognition mechanisms in a guppy population. Mol Ecol. 16:3938–3946.

Herbinger CM. 2005. Pedigree help manual. Available from: URL http://herbinger.biology.dal.ca.5080/HELP/PedigreeManual.pdf.

Herbinger CM, O'Reilly PT, Verspoor E. 2006. Unravelling first-generation pedigrees in wild endangered salmon populations using molecular genetic markers. Mol Ecol. 15:2261–2275.

Jones AG. 2005. GERUD 2.0: a computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parentage. Mol Ecol Notes. 5:708–711.

Jones B, Grossman GD, Walsh DCI, Porter BA, Avise JC, Fiumera AC. 2007. Estimating differential reproductive success from nests of related individuals, with application to a study of the mottled sculpin, *Cottus bairdi*. Genetics. 176:2427–2439.

Llaurens V, Castric V, Austerlitz F, Vekemans X. 2008. High paternal diversity in the self-incompatible herb *Arabidopsis halleri* despite clonal reproduction and spatially restricted pollen dispersal. Mol Ecol. 17:1577–1588.

Mackiewicz M, Porter BA, Dakin EE, Avise JC. 2005. Cuckoldry rates in the Molly Miller (*Scartella cristata*; Blenniidae), a hole-nesting marine fish with alternative reproductive tactics. Mar Biol. 148:213–221.

Mäkinen T, Panova M, André C. 2007. High levels of multiple paternity in *Littorina saxialis*: hedging the bets? J Hered. 98:705–711.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol. 7:639–655.

Meagher TR. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. I. Identification of most-likely male parents. Am Nat. 128:199–215.

Myers EM, Zamudio KR. 2004. Multiple paternity in an aggregate breeding amphibian: the effect of reproductive skew on estimates of male reproductive success. Mol Ecol. 13:1951–1963.

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

Neff BD, Pitcher TE, Ramnarine IW. 2008. Inter-population variation in multiple paternity and reproductive skew in the guppy. Mol Ecol. 17:2975–2984.

Neff BD, Pitcher TE, Repka J. 2002. A Bayesian model for assessing the frequency of multiple mating in nature. J Hered. 93:406–414.

Portnoy DS, Piercy AN, Musick JA, Burgess GH, Graves JE. 2007. Genetic polyandry and sexual conflict in the sandbar shark, *Carcharhinus plumbeus*, in the western North Atlantic and Gulf of Mexico. Mol Ecol. 16: 187–197.

Reusch TB. 2000. Pollination in the marine realm: microsatellites reveal high outcrossing rates and multiple paternity in eelgrass *Zostera marina*. Heredity. 85:459–464.

Sefc KM, Mattersdorfer K, Sturmbauer C, Koblmüller S. 2008. High frequency of multiple paternity in broods of a socially monogamous cichlid fish with biparental brood care. Mol Ecol. 17:2531–2543.

Simmons LW, Beveridge M, Kennington WJ. 2007. Polyandry in the wild: temporal changes in female mating frequency and sperm competition in natural populations of the tettigoniid *Requena verticalis*. Mol Ecol. 16:4613–4623.

Smith BR, Herbinger CM, Merry HR. 2001. Accurate partition of individuals into full-sib families from genetic data without parental information. Genetics. 158:1329–1338.

Song SD, Drew RAI, Hughes JM. 2007. Multiple paternity in a natural population of a wild tobacco fly, *Bactrocera cacuminata* (Diptera: Tephritidae), assessed by microsatellite DNA markers. Mol Ecol. 16:2353–2361.

Teixera S, Bernasconi G. 2007. High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. Mol Ecol. 16:4370–4379.

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

Wang J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics. 166:1963–1979.

Wang J. 2007. Parentage and sibship exclusions: higher statistical power with more family members. Heredity. 99:205–217.

Wilson AB, Martin-Smith KM. 2007. Genetic monogamy despite social promiscuity in the pot-bellied seahorse (*Hippocampus abdominalis*). Mol Ecol. 16:2345–2352.

Received May 20, 2008; Revised September 24, 2008; Accepted October 2, 2008

Corresponding Editor: Howard Ross