



Research article

Estimating the proportion of offspring attributable to candidate adults

ANTHONY C. FIUMERA,^{1,†} J. ANDREW DEWOODY,^{1,‡} MARJORIE A. ASMUSSEN¹ and JOHN C. AVISE^{1,*}

¹Department of Genetics, University of Georgia, Athens, GA 30602; Present address: [†]Department of Molecular Genetics and Biology, Cornell University, Ithaca, NY 14853; [‡]Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA
(*author for correspondence, tel.: +1-706-5421456; fax: +1-706-5423910;
e-mail: avise@arches.uga.edu)

Received 17 January 2002; accepted 3 September 2002

Co-ordinating editor: L.D. Hurst

Abstract. Statistical methods for estimating genetic parentage are increasingly applied to accommodate limited marker polymorphism and the incomplete sampling of individuals. Neff *et al.* (2000a, *Mol. Ecol.* **9**, 515–528; 2000b, *Mol. Ecol.* **9**, 529–539) published a method (*Pat*) that estimates the proportion of next-generation individuals sired by a focal male, taking into account that the male may be genetically compatible, by random chance, with offspring that are not his own. Here we employ this method to reestimate paternity of 68 nest-guarding males from several fish species. The difference between the conventional exclusion-based estimate and *Pat* was >0.05 in only four of the 68 (5.9%) fish nests analyzed. An analytical formula shows that the difference between the two estimates is expected to be negligible if the focal male is consistent with a large proportion of the genotyped offspring, or if marker polymorphism is high. In addition, computer simulations illustrate how numbers of marker loci and their levels of genetic polymorphism, as well as the mating system of the organism under study, can influence estimates of paternity derived from exclusion-based estimates and *Pat*. Finally, we discuss various applications of these estimators including cases where additional biological information is present in the form of behavioral observations on parental care.

Key words: genetic parentage analysis, maternity, mating systems, paternity, reproductive success

Introduction

Genetic studies of parentage in natural populations can become increasingly sophisticated as molecular technologies improve. It is now quite feasible to genotype hundreds or thousands of progeny and associated putative parents, allowing researchers to address previously intractable questions about maternity and paternity in nature (Hughes, 1998; Birkhead, 2000; Avise, 2001; Avise *et al.*, 2002). Associated with this effort is the need to develop appropriate

statistical estimators. The simplest of these, parentage exclusion, has been used extensively to document extra-pair fertilizations and intra-specific brood parasitism in avian taxa (see Møller and Ninni, 1998). A genetic exclusion occurs whenever the genotype of a candidate parent is inconsistent with its being the true biological parent of an offspring in question.

A common situation arises where the goal is to determine the proportion of offspring in a nest attributable to a focal putative parent. When several offspring are present, a conventional estimate of paternity (or maternity) is calculated by subtracting from 1.0 the proportion of offspring whose genotypes exclude a particular candidate father (or mother). Recently, Neff *et al.* (2000a,b) published a method that corrects this estimate by incorporating the possibility that a focal parent might also be genetically compatible with a given offspring by random chance (given the allele frequencies in the adult population). Neff's approach should be beneficial to parentage studies where large numbers of offspring have been sampled.

Although similar to parentage assignment techniques (e.g., Marshall *et al.*, 1998), Neff's method attempts to infer parentage for a focal parent or parent-pair, rather than simultaneously assessing parentage among all candidates. Neff's formulas were derived to estimate the paternity of a focal male (*Pat*) or the maternity of a focal female (*Mat*) under three biological models, one of which (the 'two-sex multiple mating model') then was used to estimate paternity of nest-guarding male bluegill sunfish (*Lepomis macrochirus*) when some of the associated fry were suspected to have been sired by cuckolding males. To be consistent with the biological setting considered here (nest-tending males), this current paper will refer to paternity estimation, but analogous results would apply to estimates of maternity in relevant situations.

The two-sex multiple mating model was developed to estimate the proportion of next-generation individuals (NGIs) parented by a given adult when multiple adults of both sexes may have genetically contributed to these offspring. The paternity of the focal male is calculated as

$$Pat = \frac{ng_{dad} - NG_{dad}}{1 - NG_{dad}}, \quad (1)$$

(Neff *et al.* 2000a), where NG_{dad} is the proportion of offspring expected to be compatible with the nest-attendant male by chance based upon allele frequencies in the population of adult breeders (equivalent to the specific exclusion probability for the neither-parent known scenario from Garber and Morris, 1983), and ng_{dad} is the observed proportion of offspring genetically compatible with the nest-attendant male. Thus, ng_{dad} is the conventional (or uncorrected) exclusion-based estimate of paternity, and *Pat* is a downward-adjusted estimate incorporating the possibility of a spurious genetic compatibility between candidate parent and juvenile. Note also that *Pat* can be

negative (unlike the traditional exclusion-based measure, which is a percentage or proportion).

Here we highlight important factors to consider when designing, implementing and interpreting parentage studies. Our goals are to: (1) estimate Pat for a large number of genetically analyzed nest-guarding males from several fish species and compare these results to published paternity estimates based on the traditional uncorrected exclusion approach; (2) analytically and numerically assess how several parameters – marker polymorphism, the proportion of offspring genetically compatible with the focal male, the genetic mating system, and the number of analyzed offspring – affect the performance of these two estimators; and (3) suggest biological scenarios where Pat may be particularly beneficial.

Methods

Paternity statistics were calculated for a total of 68 fish nests previously surveyed genetically from the redbreast sunfish (*Lepomis auritus*), the spotted sunfish (*Lepomis punctatus*), and the tessellated darter (*Etheostoma olmstedi*). Background information on these empirical studies is given in Table 1. In each of these species, breeding males guard nests into which multiple females contribute eggs (which may be fertilized by the guardian male or cuckolding males). The estimated paternity of the attendant male was calculated for each nest as: (1) the traditional exclusion-based estimate (ng_{dad}) from the published studies; and (2) the corrected Pat estimate (Neff *et al.*, 2000a). The difference between the two estimates was calculated as $ng_{dad} - Pat$. Data from three nests were disregarded in the current analysis either because the nest-guarding male was not captured, or his genotype was inconsistent with all of the analyzed NGIs from his nest.

Computer simulated nests were used to assess the performance of the exclusion-based estimate and Pat . Briefly, a specified number of adults of each

Table 1. Empirical data sets reanalyzed in the current study

Species	Number of loci	Number of nests surveyed	Mean number of embryos sampled	p_e^a	Reference
Redbreast sunfish	2	24	40	0.90	DeWoody <i>et al.</i> (1998)
Spotted sunfish	2	28	48	0.88	DeWoody <i>et al.</i> (2000a)
Tessellated darter	6	16	38	0.71	DeWoody <i>et al.</i> (2000b)

^aCumulative exclusion probability (combined across loci) calculated using Equation (2a) from Jamieson and Taylor (1997).

sex (see below) was sampled randomly from a hypothetical population at Hardy–Weinberg equilibrium, and one of those selected individuals was assigned as the focal male (i.e., the male whose paternity is to be estimated). Each selected female was equally likely to be the mother of a given offspring (NGI), and any male could mate with any female. In each trial (simulated nest), genotypes of 100 NGIs then were generated under Mendelian inheritance, arbitrarily assuming that the focal male had a probability of 0.5 of being the sire of each NGI and that the other selected males were all equally likely to have sired the remaining offspring. To avoid stochastic effects of limited sampling of progeny from a nest, we analyzed all of the 100 NGIs in each simulation trial (but see below). From the simulated genetic data, the paternity of each focal male was calculated as both the exclusion-based estimate (ng_{dad}) and Pat . In each nest, the focal male's true paternity was known, and the intent was to examine the performance of the paternity estimators to recover this known parameter.

To assess the effects of parental numbers, simulations were conducted for 10,000 nests under each of six different scenarios of contributing parents: two males (m)/2 females (f), 3m/3f, 4m/4f, 5m/5f, 10m/10f, or 'RUG' (random union of gametes). To clarify, 3m/3f indicates that the focal male had a 50% chance of being the father with the remaining two males each having a 25% chance, and each of the three females had a probability of 1/3 of being the mother (independent of which male was the father). Under RUG, the focal male again had a probability of 0.5 of being the father of any given offspring, but the offspring's other allele (and both alleles for offspring not sired by the focal male) were randomly chosen based on the allele frequencies in the adult population. RUG was designed to approximate conditions under which a large (effectively infinite) number of parents contributed to the simulated nest.

A single locus with 35 equally frequent alleles was used in these simulations so that the proportion of offspring expected to be compatible with the focal male by chance (average $NG_{\text{dad}} = 0.11$) was similar to the empirical value observed in the actual fish nests surveyed from nature (recalculated from published data as average $NG_{\text{dad}} = 0.12$). To assess the effects of marker polymorphism (NG_{dad}) on the exclusion-based estimate and Pat , simulations were conducted for 10,000 trials under 18 other scenarios that differed in the numbers of equally frequent alleles at one or multiple loci (Table 2). In these trials, five parents of each sex contributed to the nest as described above. The effect of varying male contribution was also investigated for a set of empirically determined allele frequencies. Here it was assumed that two males sire the progeny with the focal male contributing between 10 and 90% of the 100 offspring analyzed, and five females contributing equally. Two loci were used, with the number of alleles and frequencies equal to those estimated in the redbreast sunfish study (see Table 1 and DeWoody *et al.*, 1998).

Table 2. Average NG_{dad} values for various numbers of loci with equally frequent alleles at each locus

	Number of equally frequent alleles					
	5 alleles	10 alleles	15 alleles	20 alleles	25 alleles	50 alleles
1 locus	0.583	0.343	0.241	0.186	0.150	0.078
2 loci	0.340	0.120	0.058	0.034	0.023	0.006
5 loci	0.068	0.005	0.0008	0.0002	7.8×10^{-5}	2.8×10^{-6}

These values can be calculated as $(1-p_e)^n$ where p_e is the exclusion probability assuming neither-parent known (Equation (2b) from Jamieson and Taylor, 1997) and n is the number of loci.

In all simulations, allele frequencies in the population were assumed to be known, rather than estimated from a large population sample. Under each set of conditions, the following were recorded: (1) the average difference between the true paternity and each estimate of paternity; (2) the percentage of trials in which each estimator ‘perfectly’ estimated paternity (i.e., was within 0.01 of the true paternity of the focal male); and (3) the average magnitude of the error when the estimator was less than perfect. The magnitude of the error was calculated as the absolute value of the difference between the true and estimated paternity.

A final set of simulations was completed to investigate the effects of sampling a limited number of offspring from the nest. A single locus with 10 equally frequent alleles was used and either 10, 25, 50 or 100 offspring were analyzed from a total nest of 100 individuals. Simulations were conducted assuming that either two males and two females (2m/2f) or 10 males and 10 females (10m/10f) contributed to the simulated nests.

Results

Empirical difference between paternity estimators

For most of the real-life fish nests analyzed, the difference between the uncorrected (ng_{dad}) and corrected (Pat) estimates of paternity was negligible (Table 3). Indeed, incorporating the possibility of a spurious compatibility of offspring with the focal male changed the traditional paternity estimate by more than 0.05 in only four of the 68 nests (5.9%) surveyed from nature (in bold in Table 3). In two of these nests (LA11 and EO8), the relative contributions of the nest-guarding males were estimated effectively as 0% by Pat (i.e., Pat was <0), even though the focal males were consistent at face value with having sired 8.0 and 23.3% of the NGIs, respectively. Additionally, Pat would be negative when a male was not compatible with any of the sampled offspring.

Table 3. Paternity estimates for nest-guarding males in empirical genetic studies of three fish species (LA, *Lepomis auritus*; LP, *L. punctatus*; EO, *Etheostoma olmstedii*)

Nest	NG_{dad}	Exclusion-based Estimate		Difference ($ng_{\text{dad}} - Pat$)
		(ng_{dad})	Pat	
LA01	0.021	0.900	0.898	0.002
LA11	0.081	0.080	0.000	0.080
LA12	0.162	0.940	0.928	0.012
LA13	0.048	0.983	0.982	0.001
LA18	0.068	0.955	0.951	0.004
LA19	0.243	0.979	0.973	0.006
LA26	0.052	0.994	0.994	0.000
LA27	0.021	0.957	0.956	0.001
LA28	0.053	0.990	0.989	0.001
LA35	0.011	0.727	0.724	0.003
LA36	0.086	0.945	0.929	0.006
LP01	0.216	0.983	0.978	0.005
LP02	0.009	0.980	0.980	0.000
LP10	0.239	0.982	0.977	0.005
LP12	0.065	0.957	0.955	0.002
LP13	0.098	0.963	0.959	0.004
LP14	0.124	0.989	0.987	0.002
LP17	0.146	0.979	0.976	0.003
LP18	0.065	0.976	0.975	0.001
LP21	0.275	0.980	0.973	0.007
LP23	0.228	0.932	0.912	0.020
LP27	0.044	0.983	0.982	0.001
LP29	0.034	0.958	0.957	0.001
EO4	0.154	0.400	0.291	0.109
EO6	0.261	0.475	0.290	0.185
EO8	0.239	0.233	0.000	0.233
EO15	0.156	0.936	0.924	0.012
41 nests ^a	0.021–0.772	1.00	1.00	0.000
Mean	0.157	0.930	0.919	0.011
Modified mean ^b	0.118	0.827	0.801	0.026

Shown for each of 68 nests are: the proportions of offspring whose genotypes are expected by random chance to be compatible with that of the nest-guarding male (NG_{dad}); the traditional exclusion-based estimate of paternity (ng_{dad}); the 'corrected' paternity estimate (Pat); and the difference between the uncorrected and corrected estimates. Nests in which the difference between the two paternity estimates >0.05 are in bold.

^aThe genotype of the nest-guardian male was consistent with paternity for all of his putative offspring.

^bModified mean was calculated without the 41 nests where the focal male was genetically compatible with all of the analyzed NGIs.

Analytical difference between estimators

Given the close agreement between the exclusion-based estimate and Pat in most (but not all) of the fish nests analyzed above, an analytical expression for their difference should help reveal the general biological conditions under

which these estimates are (or are not) likely to differ substantially. Subtracting Pat (Equation 1) from ng_{dad} shows that the difference between the conventional exclusion-based estimate of paternity (ng_{dad}) and Pat is

$$ng_{dad} - Pat = \frac{NG_{dad}(1 - ng_{dad})}{1 - NG_{dad}}. \quad (2)$$

Neff *et al.* (2000a,b) did not provide this explicit formula, but they correctly noted that the difference between the two estimates increases as ng_{dad} decreases or as NG_{dad} increases. Inspection of Equation (2) also reveals that the difference between the two estimates approaches infinity as NG_{dad} approaches 1, and that it does so at an increasing rate as ng_{dad} approaches 0. Two other points about Equation (2) are highlighted in Figure 1. First, if the focal male is consistent with a large proportion of the NGIs (i.e., ng_{dad} is high), then the difference between the exclusion-based estimate and Pat is very small. This is true even when NG_{dad} is rather high (as would be true if marker polymorphism is relatively low). The results from the empirically sampled fish nests further demonstrate this point. The four nests in which the difference between the two estimates differed by more than 0.05 (LA11, EO4, EO6, and EO8; Table 3) all had the focal male compatible with <50% of the analyzed offspring. Second, the difference between the two estimates is also very small when NG_{dad} is low (as would be true for highly polymorphic markers), even when the focal male is genetically compatible with only a small portion of the NGIs (i.e., when ng_{dad} is low).

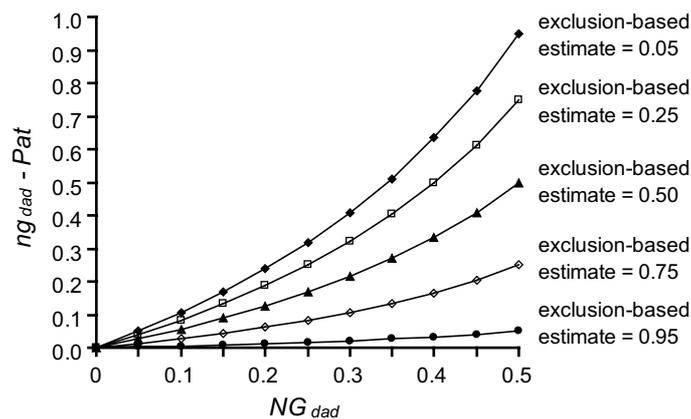


Figure 1. Analytical differences between the conventional exclusion-based estimate (ng_{dad}) and the corrected (Pat) estimate of paternity as functions of the expected proportions of offspring genetically compatible with the focal male by chance (NG_{dad}), and the proportions of offspring whose genotypes at face value are consistent with fatherhood by the focal male (ng_{dad} , exclusion-based estimate). Shown are curves (based on Equation 2) for levels of ng_{dad} ranging from 0.05 to 0.95.

Parental numbers and marker polymorphism

Computer trials (using simulated nests with known paternity) allowed us to examine how variation in parental numbers might affect the relative performance of the paternity estimators. The first point to emerge from these simulations is that under all the conditions investigated, *Pat* was an unbiased estimator of paternity. The average difference between the true value and *Pat* was virtually zero, with roughly equal numbers of estimates being too large or too small (Fig. 2). In contrast, the traditional exclusion-based estimate was biased, overestimating the true paternity of the focal male when it erred (Fig. 2). As expected, increasing the number of equally frequent alleles or the number of loci decreased the bias associated with the conventional exclusion-based method. For example, with five parents of each sex, the exclusion-based method overestimated the true paternity on average by 0.3 using a single locus with only five equally frequent alleles, by 0.1 with 20 alleles at a single locus, and by only 0.001 with five loci each with 20 alleles.

Although lack of bias is arguably the most desirable characteristic of an estimator, the average error by itself fails to convey some valuable information regarding the performance of *Pat* and the traditional exclusion-based estimate. For this reason we additionally present how often each estimator 'perfectly' estimates paternity (i.e., was within 0.01 of the true value) and the magnitude of the error when the estimator performs less than perfectly. Our criteria of 'perfect,' although arbitrary, does reveal additional information regarding the performance of these two methods not disclosed by the average error, and

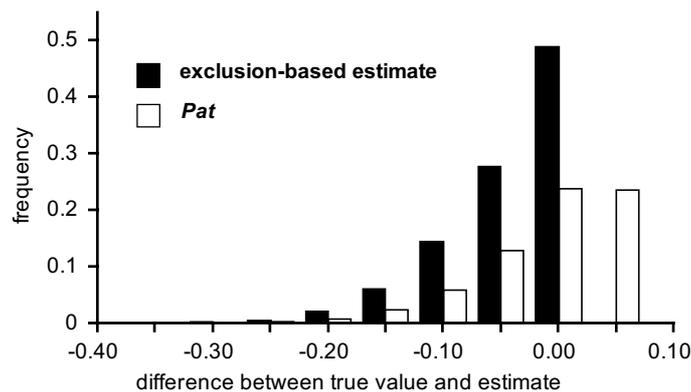


Figure 2. Histogram showing the difference between the true paternity and the estimated paternity for either the exclusion-based estimate (ng_{dad}) or *Pat* based upon 10,000 simulated fish nests. Results shown assume a single locus with 35 equally frequent alleles and five parents of each sex contributing to the progeny array.

it provides an intuitive criteria by which to assess the variability of the error estimates.

The number of contributing parents clearly influences these estimators, as demonstrated by how often the true value is estimated perfectly (Fig. 3A). The percentage of trials where the exclusion-based estimate was within 1% of the true value declined steadily as the number of parents contributing to the NGIs in a nest increased, whereas an opposite trend was observed for the *Pat* estimator. In addition, the number of parents contributing to the NGIs also influenced the magnitude of error when the estimator was less than perfect (Fig. 3B). For both estimators, this absolute error decreased as the number of contributing parents increased, especially for the exclusion-based estimate. Although *Pat* was less likely to estimate paternity perfectly, its average absolute error was smaller than for the exclusion-based estimate.

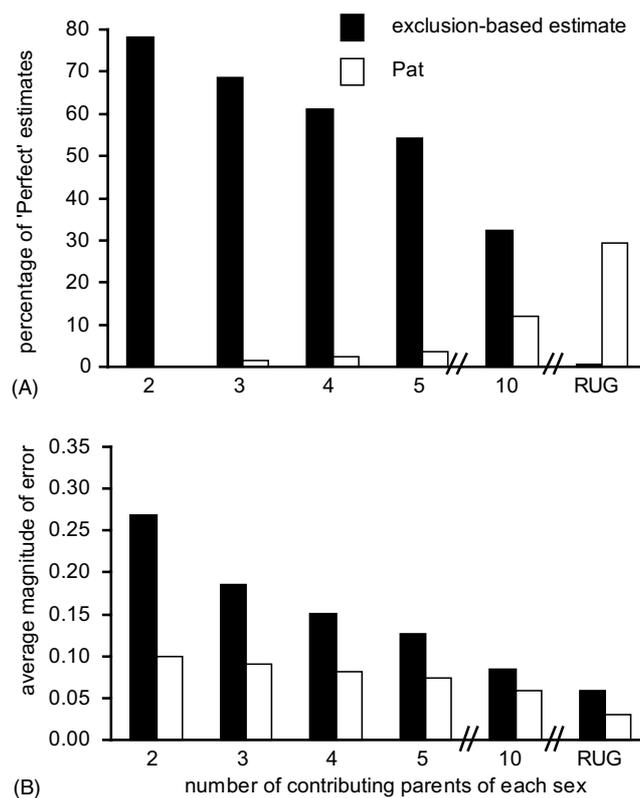


Figure 3. Percentages of estimates (from a total of 60,000 simulated fish nests) in which either *Pat* or the conventional exclusion-based estimate of paternity (ng_{dad}) were: (A) perfect, and (B) imperfect. The 'perfect' estimates were those within 1% of the true (computer-known) value of paternity. Shown are the outcomes under six different scenarios for parental numbers per nest, assuming a single locus with 35 equally frequent alleles (average $NG_{\text{dad}} = 0.110$).

Results described thus far assumed a single locus with 35 equally frequent alleles. Regardless of the number of alleles or the number of loci analyzed, *Pat* appeared to be an unbiased estimator whereas the traditional exclusion-based approach overestimated paternity when it erred. The level of marker polymorphism as well as the number of loci analyzed, however, can affect the proportion of perfect estimates and the magnitude of these errors. As might be expected, the proportion of perfect estimates increased with increases in the number of loci employed or the number of equally frequent alleles at each locus (Fig. 4A–C). With one or two loci, the absolute error for both estimators generally decreased as marker polymorphism increased (Fig. 4D,E). When the number of loci was increased to five, however, *Pat*'s absolute error increased with increasing marker polymorphism (Fig. 4F), but this reflects <10% of the trials because *Pat* was perfect more than 90% of the time under these condi-

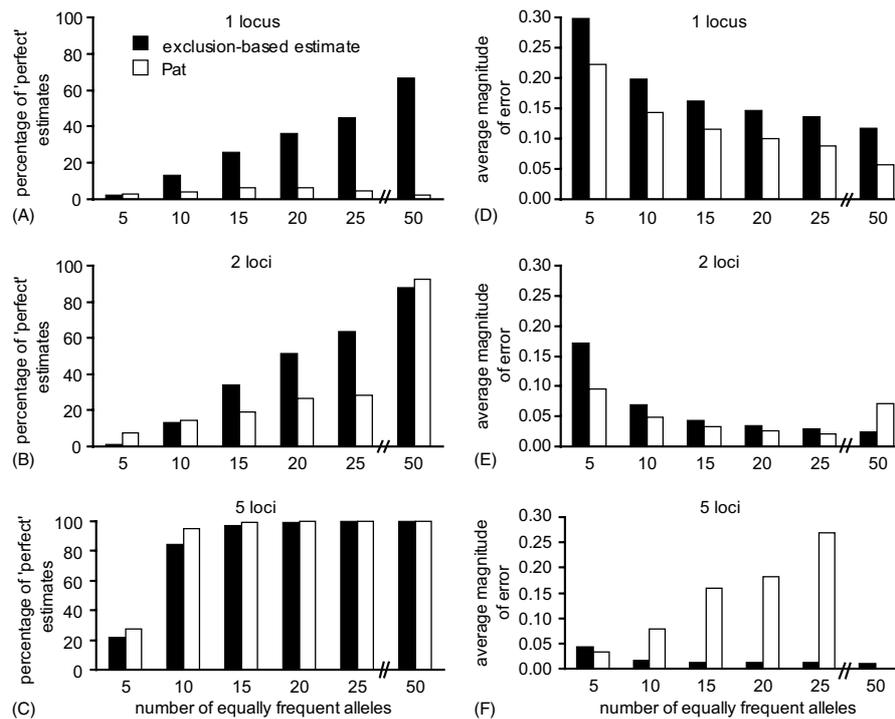


Figure 4. Percentages of estimates in which either *Pat* or the conventional exclusion-based estimate of paternity (ng_{dad}) were: (A–C) perfect, and (D–F) imperfect. ‘Perfect’ estimates were those within 1% of the true (computer-known) value of paternity. Shown are the outcomes for 18 different scenarios (a total of 180,000 simulated fish nests) that differ in the number of equally frequent alleles and in the number of loci (A and D, one locus; B and E, two loci; C and F, five loci), when five different parents of each sex contributed to the nest.

tions. Again, as compared to the exclusion-based estimate, *Pat* was less likely to estimate paternity perfectly, but when it did err the magnitude of the departure was often smaller.

It is also important to recognize that the exact performance of either estimator cannot necessarily be predicted simply by the value of NG_{dad} for the empirical markers. For example, the proportion of perfect estimates varied dramatically among three sets of markers with roughly equivalent mean NG_{dad} values (Fig. 5). Under these three conditions, the exclusion-based estimate performed best with only a single, highly polymorphic locus, whereas *Pat* performed best with multiple markers each with relatively low polymorphism.

The proportions of analyzed offspring sired by the focal male will impact the performance of both *Pat* and the conventional exclusion-based method. For both methods, the percentage of perfect estimates increased (Fig. 6A) and the average magnitude of error decreased (Fig. 6B) as the focal male truly sired a higher proportion of the analyzed offspring. As before, the conventional exclusion-based estimate tended to assess paternity perfectly more often than *Pat* (Fig. 6A), but the average magnitude of error was larger for the exclusion-based approach (Fig. 6B). These simulations also support the analytical results that the difference between the conventional exclusion-based estimate and *Pat* is small when the focal male is the true parent of a large proportion of the progeny (Fig. 6).

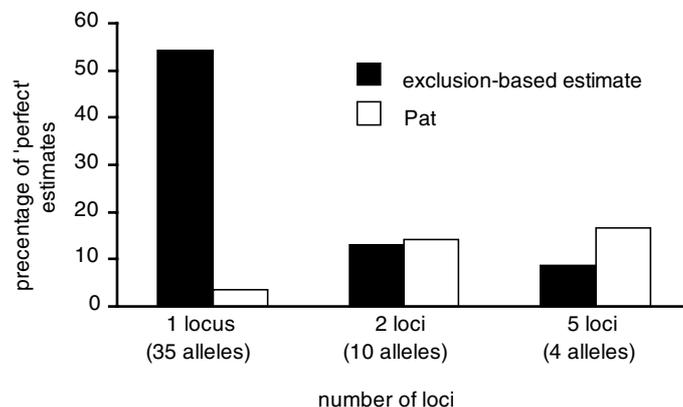


Figure 5. Percentages of estimates (from a total of 30,000 simulated fish nests) in which either *Pat* or the conventional exclusion-based estimate of paternity (ng_{dad}) were within 1% of the true (computer-known) value of paternity by the guardian male. Shown are the outcomes for three different scenarios that differ in the number of loci and in the number of equally frequent alleles, assuming that five different parents of each sex contributed to the nest. Average NG_{dad} values were chosen to be similar across the three conditions: one locus with 35 alleles (average $NG_{\text{dad}} = 0.11$), two loci with 10 alleles each (average $NG_{\text{dad}} = 0.12$), or five loci with four alleles each (average $NG_{\text{dad}} = 0.14$).

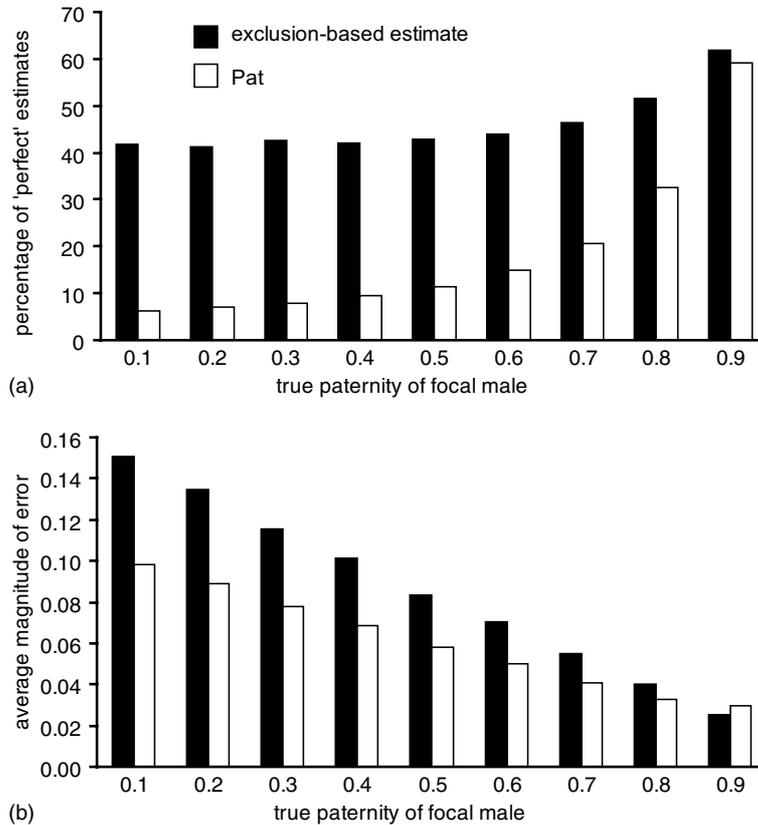


Figure 6. Percentages of estimates in which either *Pat* or the conventional exclusion-based estimate of paternity (ng_{dad}) were: (A) perfect, and (B) imperfect, assuming the empirically estimated number of loci and allele frequencies in a redbreast sunfish population (DeWoody *et al.*, 1998). 'Perfect' estimates were those within 1% of the true (computer-known) value of paternity. Shown are the outcomes for nine different scenarios (a total of 90,000 simulated fish nests) that differ in the true paternity of the focal male.

The simulations assumed that all 100 of the available offspring had been analyzed. The number of offspring sampled from the nest, however, also will affect the estimates of paternity. Regardless of the number of offspring sampled, *Pat* remained an unbiased estimator whereas the traditional exclusion-based method tended to overestimate the true paternity. The variance in the estimates, however, was affected by both the number of offspring sampled from the nest and the number of contributing parents. Although *Pat* had a larger variance than the traditional exclusion-based estimate, the variance for both estimators declined as either the number of analyzed progeny or the number of contributing parents increased (Fig. 7).

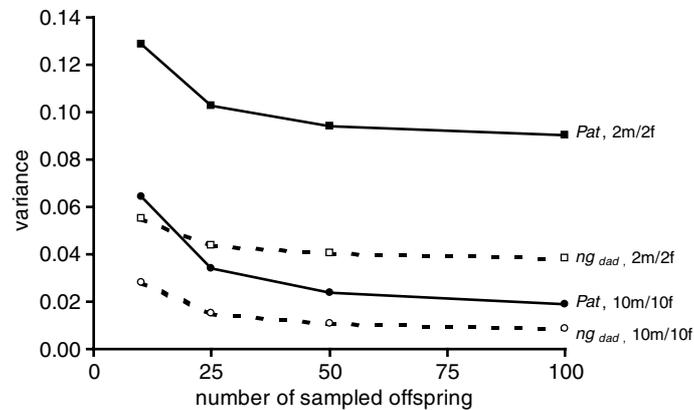


Figure 7. Variances of *Pat* and exclusion-based (ng_{dad}) estimates of paternity (from a total of 40,000 simulated fish nests) as a function of the number of offspring sampled and the number of parents contributing to the nest. For these simulations, 10, 25, 50 or 100 offspring were analyzed assuming either two males and two females (2m/2f, shown in squares) or 10 males and 10 females (10m/10f; circles) contributed to the progeny array. Estimates from *Pat* are shown with a solid line and solid marks; estimates from the exclusion-based method are shown with a dashed line and open marks. All simulations assumed a single locus with 10 equally frequent alleles.

Discussion

Due to the joint influences of natural selection and stochastic processes, most natural populations are apt to be characterized by high variances in reproductive success (e.g., Li and Hedgecock, 1998). In studying such phenomena, researchers could benefit from accurate genetic estimates of the proportion of NGIs contributed by focal adults. Toward that end, Neff *et al.* (2000a,b) provided a novel statistical approach designed to estimate genetic paternity (*Pat*) or maternity (*Mat*) in populations where offspring and their candidate parents are sampled incompletely. Here, using empirical genetic data from several nest-tending fish species, as well as analytical treatments and computer simulations, we have examined the performance of the Neff estimators vis-à-vis traditional exclusion-based estimates of genetic parentage.

Does correcting for random compatibility matter?

As demonstrated in both the theoretical and empirical appraisals, the difference between *Pat* and the conventional exclusion-based estimate of paternity can be negligible, particularly when highly polymorphic molecular markers are used or when the focal male is genetically compatible with the majority of the analyzed NGIs. Our findings are consistent with a recent report in which Neff (2001) found that the difference between *Pat* and the exclusion-based estimate of paternity was >0.05 in only two of 39 (5%) of bluegill sunfish (*Lepomis*

macrochirus) nests, that the mean difference was only 0.014, and that the exclusion-based estimate in all cases was well within the 95% confidence limits for *Pat*. Although Neff (2001) concluded that *Pat* is the most appropriate estimator, these examples also underline the point that traditional appraisals of parentage (when marker polymorphism is high) provide accurate, face-value estimates that typically differ very little from 'corrected' estimates (see also DeWoody *et al.*, 2000c; Fiumera *et al.*, 2001).

Behavioral observations and parentage

Genetic deductions can often be improved by incorporating relevant behavioral or ecological information (Hughes, 1998). Consider the two-sex multiple mating model (Neff *et al.*, 2000a). Based on its derivation, this model would seem to apply to the case in which one attempts to estimate the proportion of NGIs sired by a nest-guarding male when multiple females deposited eggs in his nest and cuckolding males may have sired a portion of the progeny. Applying this method, however, fails to acknowledge the biological observation that the focal male was captured while caring for the offspring analyzed. Thus, even if the guardian was the true sire, the *Pat* approach would attribute some fraction of those offspring to cuckolded males because of random-chance genotypic compatibility. More generally, the *a priori* probability that the guardian male is the true sire is likely neither to be one (as assumed by the exclusion-based estimate) nor equally distributed across all males in the population (as assumed by Neff *et al.*, 2000a; see also Smouse and Meagher, 1994 and Harshman and Clark, 1998). If, however, this *a priori* probability could be calculated appropriately, the derivation of *Pat* would allow it to be incorporated into the paternity estimate (see Equations A1.28 and A1.29 in Neff *et al.*, 2000b).

Appropriate biological scenarios for Pat

There are many biological settings for which the *Pat* method will likely find useful and appropriate application. For example, many fish species and marine invertebrates are broadcast spawners, release their eggs over a wide area, and provide no parental care to the young. In such contexts, the *Pat*-estimated proportion of NGIs attributable to a randomly captured adult could find proper use in estimating the reproductive success of particular potential breeders. Such data then might be applied to estimate the variance among adults in reproductive success (as required in turn, for example, for estimating effective population size), or used as a basis for identifying phenotypic correlates of fitness. In such settings, there may be no independent information (e.g., from microspatial data) on the association of offspring and probable parents

and, thus, no biological reason to suspect that one individual is more likely than another to be the parent of a randomly chosen NGI of compatible genotype. Then, correcting for random genetic compatibility could be quite valuable, as would determining the confidence limits to place around these estimates.

Recall that *Pat* performed better when many parents contributed to the analyzed offspring. When only one cuckold male contributes to the NGIs, it is unlikely that he will share any alleles with the focal male, and the traditional exclusion-based estimate will perform ‘perfectly’ for most nests. If many cuckold males were sires, however, it is more likely that some of these will happen to share alleles with the focal male. Then, the actual proportion of NGIs compatible with the focal male by random-chance approaches NG_{dad} , and the estimate of paternity calculated from *Pat* approaches the true paternity. In many broadcast-spawning fishes, tens or hundreds of individuals of both sexes often spawn in a restricted area. This is the type of condition where *Pat* appears to perform especially well in correcting for spurious attributions of genetic paternity to a focal male.

In addition to such single population applications, comparative studies could be facilitated by the Neff *et al.* (2000a,b) method. For example, *Pat* will allow researchers to compare the average paternity of males in two or more populations (of the same or different species) that differ in biologically interesting ways, such as in their operational sex ratios or population densities. The average estimate of paternity, using *Pat*, should be unbiased for large samples of focal males, thus permitting meaningful cross-population comparisons that take into account any differences in the resolving power of the genetic markers. A related aspect of the *Pat* approach is its utility in assessing the power of the available markers and in determining how to gain the most information by manipulating the numbers of individuals sampled and loci analyzed (details in Neff *et al.*, 2000b).

Additional points

Even when applied in appropriate biological settings, any statistical estimator of paternity can have limitations. Consider, for example, a situation in which many males have contributed to the analyzed NGIs in such a way that the relative contribution of any one male is likely to be small. Visual inspection of Equation (1) reveals that if $ng_{\text{dad}} \leq NG_{\text{dad}}$, then *Pat* is zero or negative. Thus, to obtain valid estimates of low-level paternity using *Pat* (or any other method), the proportions of NGIs that are genetically compatible with the focal parent by random chance must be low.

In addition, the goals for each particular study should be carefully considered, as the relative performance of various estimators may change over the

possible parameter space. For example, is the research goal to obtain unbiased population-level paternity estimates averaged across many nests, or is it to accurately estimate the number of unshared parents contributing to individual nests? Fiumera *et al.* (2001) demonstrated that a few, highly polymorphic markers performed much better than several relatively low polymorphism markers when estimating the number of parents contributing to half-sib progeny arrays. Here, however, we demonstrate that *Pat* appears to perform better when estimating paternity rates (all else being equal) with multiple loci. Although techniques such as *Pat* will likely improve paternity estimates under many biological scenarios, it is equally important to recognize that there is no entirely satisfactory statistical remedy when there is a paucity of marker polymorphism (and, conversely, no great difference among any of the paternity estimators when marker polymorphism is very high).

Acknowledgements

Work was supported by an NIH Training Grant in Genetics (to ACF), the Pew Foundation (to JCA), and NSF DEB-9906462 (to MAA). We thank R. Cartwright, M. Mackiewicz, T. Marshall, B. McCoy, D. Pearse, B. Porter, D. Walker, two anonymous reviewers and, in particular, M. Dijkstra, and B. Neff for useful comments on the manuscript.

References

- Avise, J.C. (ed.) (2001) DNA-based profiling of mating systems and reproductive ecology in poikilothermic vertebrates. *J. Hered.* **92**, 99–211.
- Avise, J.C. *et al.* (2002) Genetic mating systems and reproductive natural histories of fishes: lessons for ecology and evolution. *Annu. Rev. Genet.* **36**, 19–45.
- Birkhead, A.R. (2000) *Promiscuity: An Evolutionary History of Sperm Competition and Sexual Conflict*. Harvard University Press, Cambridge, MA.
- DeWoody, J.A., Fletcher, D.E., Mackiewicz, M., Wilkins, S.D. and Avise, J.C. (2000a) The genetic mating system of spotted sunfish (*Lepomis punctatus*): mate number and the influence of male reproductive parasites. *Mol. Ecol.* **9**, 2119–2129.
- DeWoody, J.A., Fletcher, D.E., Wilkins, S.D. and Avise, J.C. (2000b) Parentage and nest guarding in the tessellated darter (*Etheostoma olmstedii*) assayed by microsatellite markers (Perciformes: Percidae). *Copeia* **2000**, 740–747.
- DeWoody, J.A., Fletcher, D.E., Wilkins, S.D., Nelson, W.S. and Avise, J.C. (1998) Molecular genetic dissection of spawning, parentage, and reproductive tactics in a population of redbreast sunfish, *Lepomis auritus*. *Evolution* **52**, 1802–1810.
- DeWoody, J.A., Walker, D. and Avise, J.C. (2000c) Genetic parentage in large clutches: theoretical estimates and empirical appraisals. *Genetics* **154**, 1907–1912.
- Fiumera, A.C., DeWoody, Y.D., DeWoody, J.A., Asmussen, M.A. and Avise, J.C. (2001) Accuracy and precision of methods to estimate the number of parents contributing to a half-sib progeny array. *J. Hered.* **92**, 120–126.

- Garber, R.A. and Morris, J.W. (1983) General equations for the average power of exclusion for genetic systems of n codominant alleles in one-parent and no-parent cases of disputed parentage. In: R.H. Walker (ed.) *Inclusion Probabilities in Parentage Testing*. American Association of Blood Banks, Arlington, VA, pp. 277–280.
- Harshman, L.G. and Clark, A.G. (1998) Inference of sperm competition from broods of field caught *Drosophila*. *Evolution* **52**, 1334–1341.
- Hughes, C. (1998) Integrating molecular techniques with field experiments in studies of social behavior: a revolution results. *Ecology* **79**, 383–399.
- Jamieson, A. and Taylor, St.C.S. (1997) Comparisons of three probability formulae for parentage exclusion. *Anim. Genet.* **28**, 397–400.
- Li, G. and Hedgecock, D. (1998) Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. *Can. J. Fish. Aquat. Sci.* **55**, 1025–1033.
- Marshall, T.C., Slate, J., Kruuk, L.E.B. and Pemberton, J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**, 639–655.
- Møller, A.P. and Ninni, P. (1998) Sperm competition and sexual selection: a meta-analysis of paternity studies of birds. *Behav. Ecol. Sociobiol.* **43**, 345–358.
- Neff, B.D. (2001) Genetic paternity analysis and breeding success in bluegill sunfish (*Lepomis macrochirus*). *J. Hered.* **92**, 111–119.
- Neff, B.D., Repka, J. and Gross, M.R. (2000a) Parentage analysis with incomplete sampling of candidate parents and offspring. *Mol. Ecol.* **9**, 515–528.
- Neff, B.D., Repka, J. and Gross, M.R. (2000b) Statistical confidence in parentage analysis with incomplete sampling: how many loci and offspring are needed? *Mol. Ecol.* **9**, 529–539.
- Smouse, P.E. and Meagher, T.R. (1994) Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (*Liliaceae*). *Genetics* **136**, 313–322.