

Genetic Paternity Analysis and Breeding Success in Bluegill Sunfish (*Lepomis macrochirus*)

B. D. Neff

Fish have some of the most complex mating systems known in the animal kingdom. With the advent of powerful genetic markers and an emerging mathematical framework to calculate parentage, it is now possible to analyze genetic relatedness and gene flow in these systems. An important example is the bluegill sunfish (*Lepomis macrochirus*) which consists of parental males that provide sole care for the young, cuckolded males that parasitize the parentals, and females that actively choose among males within dense breeding colonies. In this article genetic markers for bluegill are characterized and their utility in parentage studies is demonstrated by calculating the genetic relatedness of parental males to their broods for an entire natural breeding colony. A novel Monte Carlo simulation is developed to calculate the confidence in the relatedness estimates and these data are used to provide an estimate of the mean breeding success of parental and cuckolded males. Finally, the applications of genetic analyses to understanding mating systems, parental care, and life-history evolution in bluegill are discussed.

Despite the enormous diversity of reproductive behaviors in fishes (e.g., Godin 1997), little is known about the complex social interactions and decisions made by individuals during mate choice and parental investment (reviewed by Andersson 1994; Birkhead and Møller 1998). Further, the mechanism underlying the evolution of alternative reproductive phenotypes is only beginning to be understood (reviewed by Gross 1984, 1996; Taborsky 1997, 1998). With the advent of molecular techniques (e.g., Avise 1994; Jarne and Lagoda 1996; Neff et al. 2000a) evolutionary and behavioral ecologists are now able to genetically quantify reproductive success in the wild and provide new insights into such processes. Of the fishes, bluegill sunfish (*Lepomis macrochirus*) is one of the most studied species (e.g., Cargnelli and Gross 1996; Dominey 1980, 1983; Gross 1980, 1982, 1991; Neff 2000) and provides a model system to apply genetic markers to research on mating system, parental care, and life-history evolution.

Bluegill are native to freshwater lakes and streams in North America, but are now found throughout much of the world (Lee et al. 1980). Bluegill exhibit one of the most social and complex mating systems in nature (Gross 1982). Males are characterized by a discrete polymorphism in life histories termed “parental” and “cuckold-

er” (Gross 1982). Parental males delay maturation and compete to construct nests in colonies, court females, and provide sole parental care for the young within their nest. By contrast, cuckolded males do not build nests of their own or care for their offspring. Instead, cuckolders mature precociously and steal fertilizations in the nests of parental males through two tactics: younger and smaller “sneakers” hide behind plants and debris near the nest edge and opportunistically dart into the nest during female egg releases; older and larger “satellites” are about the size of mature females and by expressing female coloration and behavior are able to deceive the parental male into perceiving that he has two females in his nest. Bigamy, in which two females release eggs simultaneously in a nest, occurs naturally about 10% of the time in Lake Opinicon (Ontario, Canada) and is the background against which mimicry has evolved (Gross 1982, 1991). Parental males readily detect and chase sneakers out of their nest, but only rarely detect and chase satellites. Cuckolders die before the age of mature parentals and never themselves become parentals. Spawning involves interactions between numerous individuals, including a parental male, multiple cuckolded males, and females, and results in several thou-

From the Department of Neurobiology and Behavior, Seeley G. Mudd Hall, Cornell University, Ithaca, NY 14853. This article reports on work that appears in the author's PhD dissertation completed while in the Department of Zoology at the University of Toronto. I greatly appreciate the contributions made by my supervisor Dr. Mart Gross. I also thank Dr. Joe Repka for discussions on the paternity analysis and Dr. Andrew DeWoody, Peng Fu, and Dr. Peter Smouse for comments on the manuscript. Address correspondence to Bryan D. Neff at the address above or e-mail: bdtneff@netscape.net. This paper was delivered at a symposium entitled “DNA-Based Profiling of Mating Systems and Reproductive Behaviors in Poikilothermic Vertebrates” sponsored by the American Genetic Association at Yale University, New Haven, CT, USA, June 17–20, 2000.

© 2001 The American Genetic Association 92:111–119

Table 1. The 11 bluegill microsatellite loci

Locus	Primer sequence 5'-3'	No. of alleles	No. of effective alleles ^a	Expected heterozygosity	Expected NG_{dad} ^b
<i>Lma20^f</i>	F: GGCACCTAATCTAATTGTAGCC R: TTGTGTCTGTCATTGGAATC	9	2.6	.61	.78
<i>Lma21^c</i>	F: CAGCTCAATAGTTCTGTCAGG R: ACTACTGCTGAAGATATTGTAG	8	2.3	.57	.82
<i>Lma87^c</i>	F: ATGACACAGACTCACCATGC R: CTCCTGCCATAAATCAGAC	5	2.4	.58	.82
<i>Lma102^d</i>	F: CTGTGAAAATGGTGTGAGCG R: AAACACAAAAGTCCACGCAC	9	3.3	.69	.72
<i>Lma113</i>	F: ACCATGAGCCAATGTTGAAC R: CATACGCACACACACGCAC	19	2.7	.62	.74
<i>Lma116^d</i>	F: GAATCTGGTGCACACTGAC R: GTATTGACCCATAGCAGCTGC	13	2.3	.56	.81
<i>Lma117^d</i>	F: TCACTCTGCAACCTACAAGCC R: AGATCTACATGCCACTCATTGC	7	3.3	.70	.72
<i>Lma120^d</i>	F: TGTCACCCAACTTAAGCC R: TAAGCCCATCCCAATTATCC	14	2.5	.60	.80
<i>Lma121^d</i>	F: CACTGTGGGAGAGGAAGAG R: AGTGATTTTCCCCTGGCAG	7	2.1	.53	.84
<i>Lma122^{d,e}</i>	F: AAATAAAGTTTGGGATACCCTG R: GTCGTGAGCTATTGCTTAAATAC	15	4.6	.78	.58
<i>Lma124^d</i>	F: CATTCAATTTGTCGCGACAC R: AAACACAAAAGCCACCATG	9	2.0	.50	.86
Average		10.4	2.7	0.61	0.77
Range		5-19	2.0-4.6	.50-.78	.58-.86

Number of alleles, number of effective alleles, expected heterozygosity, and expected NG_{dad} were calculated from a sample of 232 presumed unrelated individuals collected from Lake Opinicon.

^a Effective number of alleles accounts for the frequency of each allele and was calculated according to Neff et al. (2000c).

^b The expected value of NG_{dad} is inversely related to the power of the locus to detect cuckoldry, and was calculated according to Neff et al. (2000c). Generally, lower NG_{dad} values provide more precise paternity results (see Neff et al. 2000c).

^c Primer sequences first reported in Colbourne et al. (1996).

^d Primer sequences first reported in Neff et al. (1999).

^e Locus has null alleles; see Neff et al. (1999).

sand embryos of mixed parentage being raised by a single parental male.

Similar alternative life histories are found in many fish species (Gross 1984), and opposing hypotheses exist to explain their evolution (Gross 1996). If the life histories represent either alternative strategies or a mixed strategy, then equal mean fitnesses for individuals displaying either life history are required to sustain the polymorphism (Gross 1996; Maynard Smith 1982; van Damme 1991). Conversely, if the life histories represent alternative tactics within a single conditional strategy, then unequal fitnesses are expected (Gross 1996; Repka and Gross 1995). Gross (1982) and Gross and Charnov (1980) develop models that test whether alternative life histories have equal fitnesses, thus providing insight into the mechanism underlying their evolution. They tested their models using behavioral estimates of breeding success in a bluegill population and demonstrated that cuckold and parental males may have equal fitnesses. However, genetic parentage analysis is needed to provide a more accurate test of their models. Here, breeding success in a bluegill colony is genetically quantified,

and these data are compared to the behavioral inference.

This article has four objectives. First, the utility of microsatellites and new genetic models for parentage analysis is demonstrated by calculating the paternity of parental male bluegill from a breeding colony in Lake Opinicon. These parentage models have not previously been employed in a large empirical study. Second, a novel Monte Carlo simulation is developed to calculate the confidence in the paternity estimates, and trade-offs between sampling numbers of offspring and loci are discussed. Third, the paternity estimates are used to provide a preliminary calculation of the mean breeding success of parental and cuckold males, and these data are compared to previous estimates based on behavioral observations during spawning. Fourth, the utility of genetic analyses in research on mating system, parental care, and life-history evolution in bluegill is briefly discussed.

Methods

The Colony

In June 1996, a bluegill colony was carefully selected to represent the many that

have been studied in Lake Opinicon. For example, the colony was of average size and depth, and had characteristic egg scores, vegetation, and substrate (see Gross 1982, 1991; Philipp and Gross 1994). Once spawning began, a large enclosure was constructed across the mouth of the bay containing the colony. The enclosure allowed all natural behaviors to occur, but prevented dispersal of the breeding individuals. The colony occupied less than 15% of the enclosure and the net did not interfere with the breeding dynamics of the colony. Divers recorded the breeding behavior, including sneaker and satellite intrusions into the nests, and the subsequent care behavior of parental males (see Neff 2000). At the end of the care period, all individuals within the enclosure were collected, including females, parental males, and cuckolders, with either dip or seine nets, and fry were collected from each nest using diving gear (SCUBA). A random sample of the fry (100-150) from each nest was preserved in 70% ethanol for later microsatellite DNA analysis. The remaining fry were dried on filter paper and weighed to provide clutch size estimates.

Paternity Analysis

Using microsatellite multiplexing methods described in Neff et al. (2000a), genotypes at up to 11 loci were obtained for all breeding adults and a random sample of fry from each nest. A description of the loci is presented in Table 1. Within the colony, the paternity of each nest-tending parental male to its brood was calculated using his multilocus genotype, the genotypes of the young in his nest, the allele frequencies within the breeding population, and the two-sex paternity model presented in Neff et al. (2000b). The two-sex model was selected since multiple females and multiple males may have spawned in each nest and the broods may have been a product of multiple mating by both sexes. The statistical confidence (95% confidence interval) in each paternity estimate was calculated using the corresponding two-sex confidence model presented in Neff et al. (2000c). Since the confidence model is computationally intense to solve when numerous loci are used (e.g., more than three), a novel Monte Carlo simulation was developed to greatly expedite the calculation. Such Monte Carlo approaches have been effectively used to estimate other important confidence statistics in parentage analysis (e.g., Bernatchez and Duchesne 2000; DeWoody et al. 2000a;

Marshall et al. 1998). The simulation developed here increases the utility of the two-sex confidence model and will facilitate its application in parentage analyses. The program is available from the author (two-sex paternity and confidence).

Generally, the Monte Carlo simulation (Figure 1) generates a distribution of the probability of observing k of C offspring that are compatible with the putative father (i.e., share at least one allele with him at each of the L loci) over the possible paternities for the putative father (i.e., $Pat = 0-100\%$). Here, Pat is defined as the proportion of the young within a brood that are the genetic offspring of the putative parent (the two-sex paternity model calculates the most likely value of Pat), C represents the number of offspring analyzed from a given brood, and k represents the number of offspring that are compatible with the putative father as calculated from the genetic data. The 95% confidence interval is then calculated by determining the paternity values that cut off the lower and upper 2.5% of the area under the distribution. This latter calculation assumes that the a priori probability distribution of paternity is uniform, and is the least biased in the absence of additional information (see Neff et al. 2000c; Neff BD, et al., in review).

Specifically, given a value of Pat , the probability that k of the C offspring are compatible with the putative father is calculated as follows. Based on the breeding population allele frequencies, genotypes are generated for F cuckolded fathers and M mothers at each of the L loci. F and M represent the effective number of breeders contributing to the brood in addition to the putative father (see Neff et al. 2000c). Genotypes for each of the C offspring are then determined by first probabilistically selecting a father based on Pat and randomly selecting one of the M mothers. If the offspring is cuckolded, with probability $1-Pat$, then one of the F cuckolded fathers is randomly selected. One of the two alleles from each father and mother is randomly assigned to the offspring at each locus. Once genotypes for the C offspring are generated, the number that shares at least one allele with the putative father at every locus is determined. Of course, all of the putative father's offspring will be compatible with him, but an additional number that were not fertilized by him may also be compatible by chance. This "chance" that another male's offspring are compatible with the putative father's genotype decreases as the number

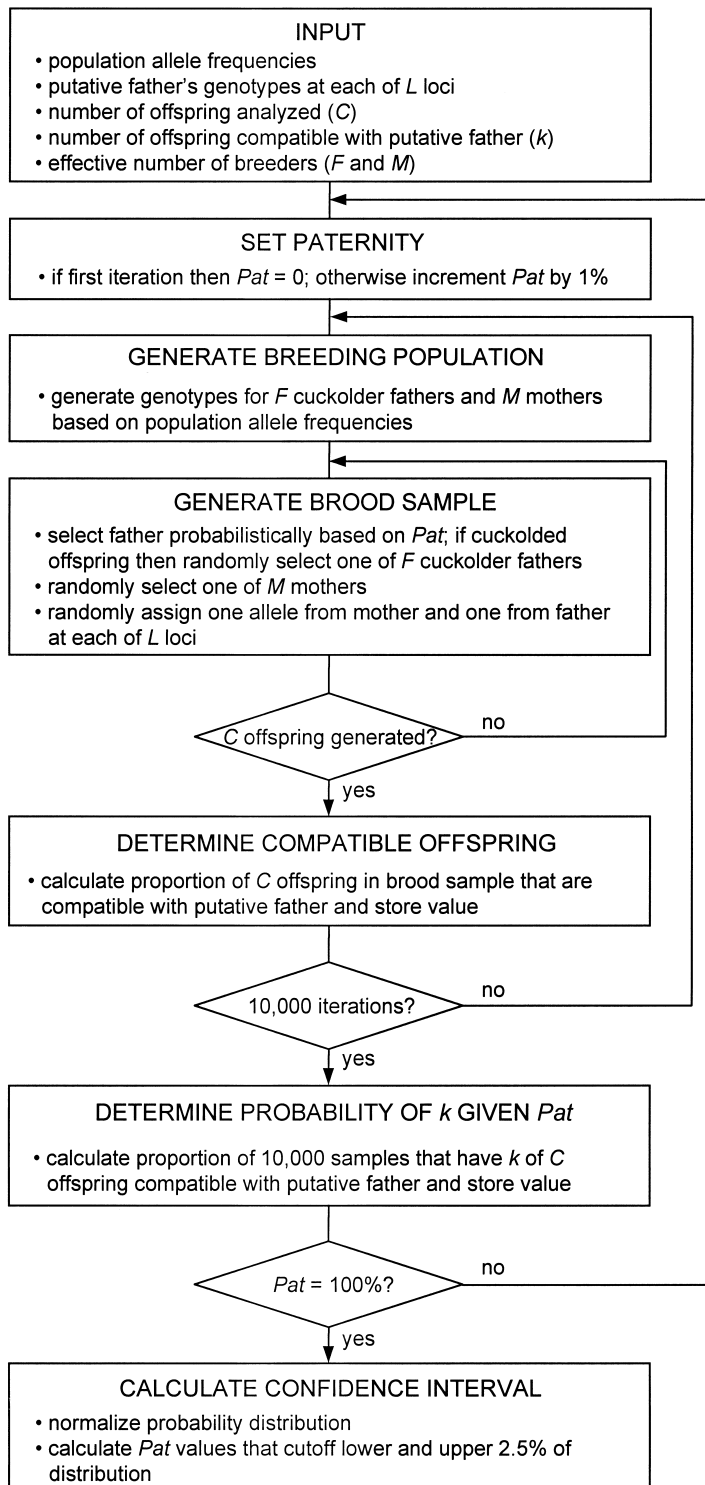


Figure 1. Schematic of the Monte Carlo simulation used to evaluate the two-sex paternity (or maternity) confidence model developed in Neff et al. (2000c). The simulation provides the paternity (or maternity) of the putative parent and the 95% confidence interval.

of loci or the frequency of the putative father's alleles decreases (see Neff et al. 2000b).

For each possible value of Pat (0–100%), the entire process is performed 10,000 times and the proportion of the samples that have k compatible offspring is deter-

mined. Finally, the distribution is normalized such that the probability that $Pat \in \{0-100\%$ is 1 and the confidence interval is calculated as above. As an example, if for a given value of Pat , very few of the 10,000 samples contained k compatible offspring, then it is unlikely that Pat re-

flects the putative father's actual paternity. Conversely, if most of the samples contained k compatible offspring, then it is likely that *Pat* is the putative father's actual paternity.

In summary, the simulation requires the following input: (1) the putative father's genotypes at each of the L loci; (2) the frequency of each allele in the breeding population at each of the L loci; (3) the number of offspring analyzed from the putative father's brood; (4) the number of offspring in the brood sample that were compatible with the putative father (k); and (5) the effective number of cuckolded males and females that have contributed to the brood. The effective number of breeders must generally be estimated. In Neff et al. (2000c) several estimation methods are discussed. Here the average population ratios of breeding females and cuckolded males to parental males was used (cuckolded males = 6, females = 4) since, on average, these should be accurate [see Neff et al. (2000c) for further discussion]. The simulation provides the following output: (1) the paternity estimate for the putative father, and (2) the 95% confidence interval.

It should be noted that the simulation does not consider germline mutations, which can result in an underestimate of the putative parent's paternity. However, even with the relatively high mutation rates observed at some microsatellites (e.g., 10^{-3}), the bias is expected to be small. For example, even when 10 loci are used in the analysis, the probability of a mutation at any of the loci is less than 1%. The simulation also assumes that the genotype frequencies within the breeding populations are in Hardy-Weinberg equilibrium.

Empirical Validation

To test the accuracy of the paternity calculations and validate the models, four analyses were conducted. First, the mean paternity of the parental males was compared to estimates based on behavioral observations of intrusions by cuckolders during spawning [presented in Gross (1982) and Gross and Charnov (1980)]. Second, the mean paternity was compared to independent colony estimates based on an alternative model using allozyme data (Philipp and Gross 1994). Third, since multiple microsatellite loci were used in this study, and each locus provides an independent estimate of paternity, the following analysis was performed for each brood. Independently for each of the n mi-

cro-satellite loci used to genotype a particular brood, the paternity of the parental male was calculated using the two-sex paternity model. The estimate based on the single locus was then compared to a second estimate based on the remaining $n - 1$ loci to ensure independence of the two estimates, and the difference was calculated. Next, an average difference based on the n comparisons was calculated for each brood. Generally the estimates based on individual loci will have considerably lower confidence (i.e., precision) than the estimates based on multiple loci, but should be equally accurate (Neff et al. 2000c). Therefore the average of the mean differences across all broods should not differ significantly from zero and should be symmetrically distributed about zero (Neff et al. 2000b). That is, the single-locus analysis should provide, on average, the same paternity estimate as the more powerful analysis based on the combined loci. Fourth, multiple linear regression was used to quantify the independent effects of paternity and NG_{dad} (defined as 1-exclusion probability in Neff et al. 2000b,c) on the precision in the paternity estimates. These data were then compared to theoretical predictions presented in Neff et al. (2000c).

The number of offspring analyzed is also expected to have an effect on the precision of the paternity estimates, in part, through sampling error (Neff et al. 2000c). However, since in this study there was low variance in the number of offspring analyzed among broods (i.e., 44–46 offspring where analyzed from most broods), it was not possible to empirically test this prediction.

Relative Fitnesses

To calculate the mean breeding success of parental and cuckolded males, the paternity estimates from the colony were combined with published data from four additional colonies (Philipp and Gross 1994). The five colonies varied in physical habitat characteristics and are representative of the colonies found in Lake Opinicon. First, since nesting parental males occasionally intrude into the nests of their neighbors, the paternity of each parental male was also calculated in each nest adjacent to his own for the colony studied here. Since neighbor intrusion rates are generally low and do not differ significantly among colonies (Gross 1980), it was assumed that this neighbor intrusion success was representative of the other four colonies. The mean breeding success of

parental males was therefore calculated as the mean paternity of parental males plus the mean paternity of neighboring parental males. Conversely the mean breeding success of cuckolded males was calculated as one minus the mean paternity of parental males minus the mean paternity of neighboring parental males. The clutch size of each nest was not included in the analysis, since clutch size estimates were not available for the four additional colonies. This likely should not bias the fitness estimates significantly, since there was no relationship between clutch size and paternity based on the colony studied here (see Neff 2000; but also see Discussion below).

The relative fitnesses of parental and cuckolded males were calculated by entering the mean breeding successes into the life-history models developed by Gross (1982) and Gross and Charnov (1980). Briefly, these models compare the proportion of eggs fertilized by each life-history group to the proportion of male individuals that enter that group. If these proportions are equivalent, then the two life histories have equal fitnesses.

All simulations were written in the C programming language and statistical analyses were performed using SPSS (version 10).

Results

Microsatellites

The microsatellite characteristics for the 11 loci are presented in Table 1. Across all loci there was an average of 10.4 alleles, but only 2.7 effective alleles. The effective number of alleles incorporates the skew in allele frequencies [see Neff et al. (2000c) for further discussion]. Of interest, allele number and effective allele number were not correlated ($r = 0.31$, $P = .18$, $n = 11$; one-tailed test), suggesting that as allele number increases so does the skew in allele frequencies. The mean heterozygosity of the loci was 0.61 and the average expected value of NG_{dad} (i.e., 1-exclusion probability) was 0.77 and ranged from 0.58 at *Lma122* to 0.86 at *Lma124*. Since confidence in the paternity estimates is inversely correlated with NG_{dad} (Neff et al. 2000c), *Lma122* was the single most informative locus. However, null alleles were detected at this locus (Neff et al. 1999), and therefore it may only be informative when the putative father is heterozygous (but see Discussion). Across all 11 loci, the cumulative expected value of NG_{dad} was 0.05. As expected, the effective num-

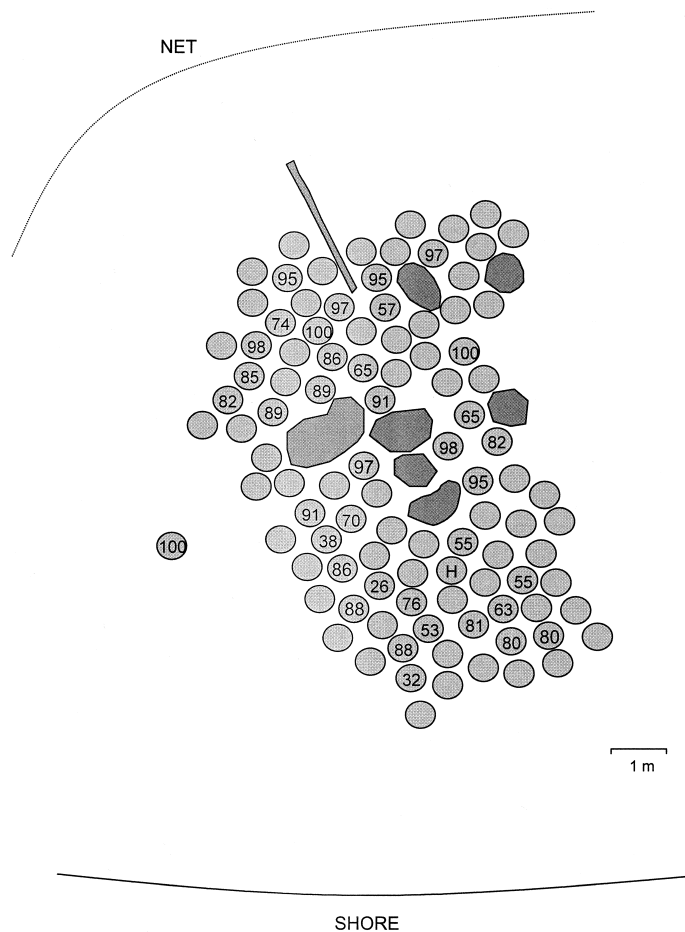


Figure 2. Map of the bluegill colony. Circles represent nests and the numbers are the paternity of the attendant parental male. Cuckoldry by neighboring parentals was rare and represented less than 2% of the total eggs fertilized. All other paternity was attributed to cuckold male. Nests without numbers represent males that deserted after spawning. Any young in these nests were lost to predation. Forty parental males were successful at recruiting fry. One nondeserter was a bluegill-pumpkinseed hybrid (marked "H"). The hybrid had a paternity of zero and was probably infertile. Dark irregular objects represent large rocks or submerged woody debris. A paternity estimate could not be made for one successful male (see text).

ber of alleles was highly correlated with the expected heterozygosity ($r = 0.98$, $P < .001$, $n = 11$; one-tailed test), and both of these variables were highly correlated with the expected value of NG_{dad} (effective alleles: $r = 0.99$, $P < .001$, $n = 11$; heterozygosity: $r = 0.97$, $P < .001$, $n = 11$; one-tailed test for both). Therefore microsatellite loci with even allele frequency distribution, and not necessarily more alleles, provide the lowest expected NG_{dad} values and may be the most informative in parentage analyses.

Colony Paternity

A total of 99 parental males and one bluegill-pumpkinseed hybrid constructed nests within the colony (Figure 2). Only 39 of the parental males and the hybrid provided parental care to the end of the care period. Most of the 60 parental males that abandoned did so immediately after spawning, probably a result of obtaining relatively few

eggs (see Neff 2000). These males had no recruitment, as eggs do not survive without parental care (Gross 1982; Neff 2000). Genotypes were obtained from each of the 40 males that provided parental care and an average of 44 fry from each of the broods (range 31–46, $n = 39$; DNA of sufficient quality could not be obtained from one bluegill brood due to fry degradation prior to preservation). In total, more than 18,000 genotypes were generated using up to 11 loci.

A summary of the paternity analysis is presented in Table 2 and Figure 3. The mean paternity of the 38 parental males was 78.9% (range 26–100%). The hybrid had a paternity of zero and was probably infertile. Including the hybrid, the mean paternity of parental males drops to 76.9%. Cuckoldry by neighboring parentals was rare, representing only 1.8% of the total eggs fertilized. Therefore, includ-

ing the hybrid, cuckolders fertilized an average of 21.3% of the fry within the colony.

Paternity Confidence

The confidence in the paternity results from the colony are presented in Table 2. Generally the results have high confidence, yielding a narrow 95% confidence interval. For most broods, the value of NG_{dad} is very low, and only a limited increase in precision could be achieved from increasing the number of loci. Conversely, most of the potential error in the estimates is attributed to sampling error introduced from analyzing only a portion of the brood (data not shown, but see Neff et al. 2000c). Therefore if one wanted to increase the precision of these estimates, then the best approach would be to increase the number of offspring analyzed.

The value of NG_{dad} for each male decreased with the number of loci used, but with diminishing returns. For most of the broods analyzed, 10 or 11 loci were used (Table 2). However, in most cases considerably fewer loci could provide estimates with similar precision. For example, an exclusion probability (i.e., $1 - NG_{dad}$) of at least 90% of its value based on all loci was obtained with only a single locus (the one with the lowest associated NG_{dad} value) for more than one-quarter of the broods, two loci for half the broods, and six loci for the remainder of the broods (Figure 4). Thus paternity estimates with high precision can be obtained with only a few loci (e.g., in many cases only one or two).

Empirical Validation

The mean paternity for each parental male, as calculated from the microsatellite loci individually, was similar to the value calculated from the loci collectively (Figure 5). The mean difference was 0.0017 ± 0.0176 (SE) and was not significantly different from zero ($t = 0.099$, $df = 38$, $P = .92$). Thus there was good agreement between the paternity estimates based on each locus individually and all loci collectively.

Multiple linear regression revealed that, as predicted, the precision of the paternity estimates increased with increasing paternity values ($\beta = 0.71$) and decreased with increasing NG_{dad} values ($\beta = -0.33$; $r^2 = 0.56$, $F_{3,38} = 23.0$, $P < .001$). Offspring number was also included in the analysis, but, as expected, had no significant effect due to the low variation among broods ($\beta = 0.71$, $P = .64$).

Table 2. Summary of paternity analysis including the number of offspring, the number of loci, NG_{dad} , ng_{dad} , paternity estimate, 95% confidence interval, and precision

Nest	$N_{\text{offspring}}$	N_{loci}	NG_{dad}^a	ng_{dad}^b	Pat (%)	95% CI	Precision (%) ^c
1	45	10	.1647	.9111 (41)	89	75–96	11
2	46	10	.1259	.6957 (32)	65	47–78	19
3	46	11	.0266	.8261 (38)	82	68–91	13
4	45	10	.0648	1.000 (45)	100	92–100	4
5	46	11	.0358	.5652 (26)	55	40–68	18
6	42	8	.0226	.8095 (34)	80	65–90	14
7	31	10	.1110	.8710 (37)	85	67–94	15
8	44	11	.0234	.9091 (40)	91	78–96	9
9	46	10	.0579	.6522 (30)	63	46–76	18
10	41	11	.1584	.7805 (32)	74	55–86	18
11	46	10	.0252	.5652 (26)	55	40–69	19
12	43	10	.0178	.8140 (35)	81	67–90	13
13	46	11	.0457	.8913 (41)	89	75–95	11
14	44	11	.0182	.9773 (43)	98	88–98	5
15	45	10	.1494	.6000 (27)	53	32–69	24
16	46	10	.0340	.8696 (40)	86	74–93	10
17	45	11	.1714	.8000 (36)	78	57–87	17
18	46	11	.0532	.8696 (40)	86	73–93	11
19	44	11	.0001	.5682 (25)	57	42–71	18
20	45	11	.0148	.8000 (36)	80	66–89	13
21	38	9	.1412	.9737 (37)	97	84–99	8
22	46	11	.0147	1.000 (46)	100	93–100	4
23	44	10	.1621	.9773 (43)	97	86–99	7
24	44	11	.0015	.8864 (39)	88	76–95	10
25	46	11	.1097	.8913 (41)	88	74–95	11
26	46	9	.0192	.9565 (44)	95	85–99	7
27	40	10	.0046	.3250 (13)	32	20–48	21
28	44	10	.0039	.7045 (31)	70	56–82	15
29	45	11	.0085	.2667 (12)	26	16–41	20
30 ^d	46	10	.0016	.0000 (00)	0	0–7	7
31	45	10	.0371	.6667 (30)	65	51–78	16
32	42	10	.0384	.9762 (41)	97	87–99	6
33	45	10	.0242	.9111 (41)	91	79–96	9
34	45	11	.1148	.9556 (43)	95	83–98	8
35	43	11	.0885	.8372 (36)	82	67–91	13
36	46	10	.0256	.9565 (44)	95	85–99	7
37	45	11	.1154	1.000 (45)	100	92–100	4
38	45	11	.0174	.9778 (44)	98	88–98	5
39	46	10	.1919	.5000 (23)	38	14–56	30
Average	44.2	10.4	.0626	.7830	76.9	—	12.5
Range	31–46	8–11	.00–.19	0.0–1.0	0–100	—	4–30

^a The value of NG_{dad} is inversely related to the power of the locus to detect cuckoldry and was calculated according to Neff et al. (2000c).

^b The value of ng_{dad} is equal to the proportion of the offspring analyzed from the brood that are compatible with the putative father (see Neff et al. 2000c). The number of compatible offspring is presented in parentheses.

^c Precision represents the range in the confidence interval expressed as a proportion of the actual paternity: (range in 95% CI)/(100% + paternity).

^d Male was a bluegill-pumpkinseed hybrid and was probable infertile.

Relative Fitnesses

The mean parental male paternity estimates for the four colonies studied by Philipp and Gross (1994) were 100%, 85.6%, 73.5%, and 41.3%. Adding an additional 1.8% to the latter three values to account for cuckoldry by neighboring parental males, and averaging with the value of 78.7% found in the present study, gives an overall mean of 76.9 for the percentage of young sired by parental males. Therefore the mean paternity by cuckolded males is 23.1%. Gross (1982) and Gross and Charnov (1980) found that 21% of age 2 males were cuckolders (95% binomial confidence interval 11–31%). These two proportions are quite close and suggest that individuals displaying either of the two life histo-

ries may have similar mean fitnesses (but see Discussion).

Discussion

Genetic Markers and Paternity Analysis

This study provides the first paternity analysis of individual parental males from a naturally breeding colony of bluegill. The two-sex paternity model developed by Neff et al. (2000b) was selected to calculate paternity in the colony for several reasons. First, the model enables parentage analysis in complex mating systems with two-sex multiple mating. Second, the model estimates the proportion of offspring fathered by a putative parent and thus provides the desired mean genetic re-

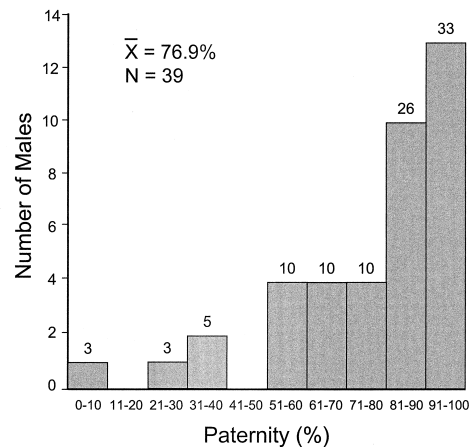


Figure 3. Histogram summarizing the paternity results for the 38 care-providing parental male bluegill and one bluegill-pumpkinseed hybrid collected from the colony. The numbers above the bars represent the percent of individuals in the corresponding category. The paternity estimates represent the most likely value of *Pat* for each parental male as calculated from the two-sex paternity model.

latedness (or paternity) of the brood to the parental male. Third, the model does not require complete sampling of the brood. This is especially important since bluegill broods, like those of many poikilotherms, are comprised of thousands of individuals. Fourth, the models can be used with many types of genetic data including microsatellites. Finally, the subset of genetic loci that have the greatest resolving power specific to each putative parent can be identified, thereby increasing the efficiency of the analysis.

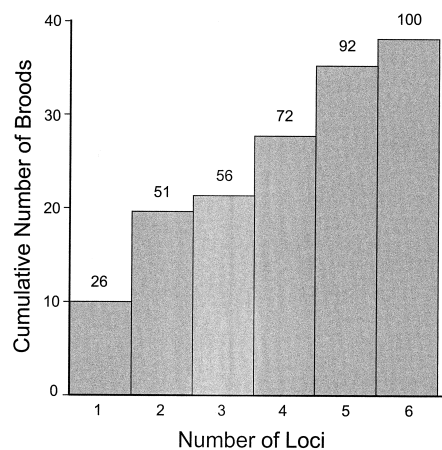


Figure 4. The number of loci required to obtain 90% of the maximum resolving power for each brood. Resolving power was calculated as $1-NG_{dad}$ and reflects the ability of the genetic analysis to detect cuckoldry (similar to an exclusion probability) and calculate paternity with high precision. Although for most broods 10 or 11 loci were employed, nearly equivalent resolving power (90%) could have been obtained for more than one-quarter of the broods with only a single locus and for half of the broods with only two loci. In no case were more than six loci required. Numbers above the bars represent the cumulative percentage of broods.

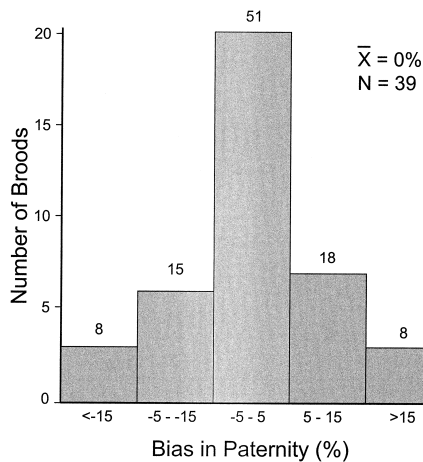


Figure 5. Histogram summarizing the comparisons of paternity estimates based on loci analyzed individually and collectively. As an example, given a brood that was analyzed with 10 loci, paternity was obtained based on each locus independently and these estimates were compared to the estimate based on the remaining 9 loci analyzed collectively. The mean difference in these 10 comparisons was calculated for the brood. As expected, across all broods the mean difference between these values (i.e., bias) was not significantly different from zero ($P = 0.92$). The distribution was also approximately symmetric about zero. Therefore the paternity inference is unbiased. Numbers above the bars represent the percentage of broods in the corresponding category.

The current genetic analysis of paternity in bluegill can be compared to previous estimates for the Lake Opinicon population. The paternity results closely match estimates based on behavioral observations during spawning (Gross 1982), as well as mean colony estimates based on protein electrophoresis (Philipp and Gross 1994). The first study observed seven colonies and reported a mean parental male paternity of 83% (range 62–97%). The latter study examined four colonies and reported a mean paternity of 75% (range 41–100%). Comparing a colony in close proximity and with similar ecology (e.g., vegetation and depth) to the colony studied here, Gross (1982) calculated a mean paternity of 84% and Philipp and Gross (1994) calculated 74%. Both estimates are close to the estimate calculated here of 76.9%. Therefore the detailed genetic analysis provided here corroborates previous estimates.

The genetic analysis of paternity also provided a means to empirically validate the two-sex paternity and confidence models. Since as many as 11 microsatellite loci were used to calculate paternity, it was possible to estimate each parental male's paternity from the loci independently as well as collectively. As predicted, the difference in these values was symmetrically distributed about zero and did not differ

significantly from zero. Therefore the loci appear to provide unbiased estimates of the parental male's actual paternity. The confidence in the paternity estimates also conformed to theoretical predictions (see Neff et al. 2000c). For example, as predicted, increased precision was associated with estimates for males that had either high paternity or a rare multilocus genotype. This study therefore provides the first empirical validation of the two-sex paternity models.

The two-sex paternity model should be effective for paternity (and maternity) analyses in many mating systems. Neff et al. (2000c) showed that precision of the paternity estimates increases as the frequency of the putative parent's genotype in the population decreases (e.g., as measured by their NG_{dad}). Generally NG_{dad} decreases with the number of loci used, but with diminishing returns. For most of the broods analyzed here, estimates with high precision could have been obtained with only a couple of the most informative microsatellite loci specific to the parental male. Based on a preliminary screening, the most informative loci specific to each male can be identified (see Neff et al. 2000b). Therefore, with only a few loci, the two-sex paternity model can provide paternity estimates with high precision, even in complex mating systems.

A particularly striking finding is that the effective number of alleles was not correlated with the actual number of alleles. This suggests that microsatellites with more alleles have a greater skew in allele frequencies. Given that many microsatellites are characterized by highly skewed allele frequency distributions (Jarne and Lagoda 1996), the lack of correlation between effective and actual allele numbers may be a common property of microsatellites. Since NG_{dad} (and the exclusion probability) is mathematically similar, and therefore highly correlated, to the effective number of alleles, and not the actual number of alleles, similar precision in parentage analysis can come from loci with fewer alleles. Therefore, given that these loci also tend to be easier to score and multiplex (e.g., Neff et al. 2000a; O'Reilly et al. 1996) researchers performing parentage analysis should consider loci with fewer alleles.

The two-sex paternity model (rather than straight exclusion methods) should be used for paternity analysis. In this case, simply calculating the paternity as the proportion of offspring that cannot be excluded (i.e., ng_{dad} ; see Table 2) might ap-

pear to provide reliable estimates. For example, the proportion of compatible offspring (ng_{dad}) and the paternity estimate calculated from the model were highly correlated ($r = 0.99$, $P < .001$, $n = 39$). However, such analysis would lead to a systematic bias, overestimating the paternity of the parental males (paired t test: $t = 20.7$, $P < .001$, $n = 39$). This may be especially important when comparing the fitnesses of parental and cuckold males. In this case, the bias would be about +2% ($[78.3-76.9]/76.9$) for the parental males and -6% ($[21.7-23.1]/23.1$) for the cuckold males. Such a bias, albeit small, could be consequential when discriminating between alternative models to explain the evolution of the life histories. Further, for some individual males, the bias can be especially high. For example, the paternity of one of the parental males (no. 39; Table 2) would be overestimated by 32% ($[50-38]/38$). This bias increases as fewer loci are used. Alternatively, if enough loci are used such that $NG_{dad} = 0$ (i.e., exclusion probability = 1), then the model is equivalent to straight exclusion. Therefore the model should always be used to ensure accurate paternity inference.

Generally microsatellite loci with null alleles can bias parentage inference and must be employed with caution in paternity analyses (e.g., Callen et al. 1993; Pemberton et al. 1995). The two-sex paternity model, however, is impervious to the effects of null alleles when the putative parent is heterozygous for known alleles and the population frequency of the null allele is known [see Brookfield (1996) for methods to estimate the frequency of null alleles]. The model adjusts the proportion of offspring in the brood that are compatible with the putative parent by the expected proportion that are compatible by chance. Given that the putative parent does not contain a null allele, neither of these proportions or the paternity estimate are affected by the null allele. Of interest, we were also able to use the two-sex paternity model to detect the null allele in homozygous parental males. For example, we found that in several cases a parental male's paternity dropped significantly when the locus with the null allele (*Lma122*) was included in the analysis, as compared to when it was not. In most other cases, however, its inclusion had little effect on the paternity estimates. Theoretically, if the putative parent had a null allele, then on average, half of the parent's offspring would inherit the null allele and could appear to be incompatible. Of

course, some of the offspring that inherit the null allele may also appear compatible with the parent when their other allele matches the parent by chance. If a putative parent is homozygous at a locus with a null allele and the parent's paternity estimate drops significantly when the locus is included in the analysis, then it is likely that the parent has the null allele. If the paternity estimate does not drop, however, in rare cases that parent may have the null allele (e.g., if the mates were all homozygous for the putative parent's visible allele). Therefore, although the presence of the null allele can generally be identified, it is not always possible to verify its absence.

Previous studies addressing statistical confidence in parentage estimates have focused on models that attempt to identify the genetic parents for each offspring (e.g., Chakraborty et al. 1974; Meagher 1986; Smouse and Meagher 1994). Confidence statistics for these models calculate the probability of identifying the true parents and the number of loci required (Chakraborty et al. 1988; Double et al. 1997; Estoup et al. 1998; Marshall et al. 1998). Models have also been developed to estimate the number of individuals contributing to a brood of half sibs and the number of offspring and loci required to provide accurate estimates (DeWoody et al. 2000a,b). These latter models calculate the level of multiple mating within a brood, but do not identify the actual parents or provide estimates of their relative success. By contrast, the Monte Carlo simulation developed here calculates the confidence interval associated with the estimated proportion of offspring fathered (or mothered) by a putative parent. It does not identify individual parent-offspring relationships or calculate the number of other parents contributing to the brood, but instead provides the minimum and maximum proportion of the offspring that the putative parent is likely to have fertilized. The simulation greatly expedites the evaluation of the two-sex confidence model and should therefore prove useful for researchers wishing to apply the model in parentage analyses.

The paternity estimates of the nest-tending parental male bluegill are in sharp contrast to genetic estimates for parental male redbreast sunfish (*Lepomis auritus*). Based on a sample of 25 broods, DeWoody et al. (1998) calculated a mean paternity of 97% (range 73–100%). Fifty-six percent (14/25) of these broods had no evidence of cuckoldry, in contrast to only the 8% (3/

39) found here. Two of the redbreast broods were associated with a parental male that likely expelled the original male sometime after the eggs had been spawned, possibly to obtain a desirable nest site (DeWoody et al. 1998). Although a single parental male within the bluegill colony (the bluegill-pumpkinseed hybrid) had a paternity of zero, he was identified prior to spawning as the original male and was instead probably infertile. Since, in bluegill, spawning generally occurs in discrete synchronous bouts, nest takeovers for the purpose of obtaining a desirable nest site are unlikely to occur after the eggs have been spawned. Although cuckoldry morphs have been documented in redbreast (e.g., Lukas and Orth 1993), DeWoody et al. (1998) did not find any cuckolders in their population sample. Further, in one of their broods they were able to attribute the cuckolded offspring to the neighboring parental male. Therefore it is possible that their population of redbreast does not contain specialized cuckolders, explaining the considerably lower cuckoldry rates as compared to the bluegill population studied here.

Genetic Analyses in Bluegill Sunfish

Genetic analyses of reproductive success in bluegill are focused at three levels: (1) the individual, (2) the colony, and (3) the population. Parentage analysis at the individual level examines the evolution of sperm competition strategies (Fu 2000; Fu P, et al., in preparation). Using genetic markers and the models developed by Neff et al. (2000b), the fertilization success of cuckolded males can be determined during individual spawning intrusions when sperm competition exists. Cuckolder success can then be correlated with phenotypic and sperm characteristics, providing insights into the evolution of sperm competition strategies in bluegill. These data will also provide a tool to accurately quantify reproductive success from spawning observations.

Parentage analysis at the colony level examines gene flow within the major social unit of bluegill reproduction (Neff 2000; Neff BD and Gross MR, in review, in preparation). These analyses quantify the phenotypic and ecological correlates of reproductive success within the breeding bout, as well as individual mate choice and parental investment decisions. For example, the variance in paternity among parental males is, in part, explained by male energy reserves, parasite load, and fluctuating asymmetry (Neff 2000; Neff BD and Gross MR, in preparation).

Parentage analysis at the population level examines the evolution of alternative reproductive life histories. This analysis can provide data needed to discriminate between opposing hypotheses: (1) alternative strategies, (2) mixed strategies, and (3) the conditional strategy (Gross 1996). The genetic calculation of the mean breeding success of cuckolded and parental male bluegill provided in this article is consistent with previous estimates based on behavioral observations and suggests that individual fitnesses under the two life histories are similar. However, a complete test would include a survey of all colonies during the breeding season, since cuckoldry rates can vary widely through the breeding season and among colonies (Gross 1980, 1982). Further, the paternity estimates should also incorporate brood sizes and the expected survivorship of each offspring to age 2 (when the decision is made to mature precociously and become a cuckolded, or to delay maturation and become a parental). Nevertheless, the calculation presented here demonstrates the application of the paternity models and their potential to address the evolution of alternative reproductive life histories in nature.

With the advent of powerful molecular genetic techniques and the emerging framework for parentage analysis, evolutionary and behavioral ecologists are now able to calculate genetic relatedness and gene flow within wild populations. These techniques are providing new and previously unattainable insight into the evolution of mating systems and life histories.

References

- Andersson M, 1994. Sexual selection. Princeton, NJ: Princeton University Press.
- Avice JC, 1994. Molecular markers, natural history and evolution. New York: Chapman & Hall.
- Bernatchez L and Duchesne P, 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Can J Fish Aquat Sci 57:1–12.
- Birkhead TR and Møller AP, eds, 1998. Sperm competition and sexual selection. San Diego, CA: Academic Press.
- Brookfield JFY, 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5:453–455.
- Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, and Sutherland GR, 1993. Incidence and origin of "null" alleles in the (AC)_n microsatellite markers. Am J Hum Genet 52:922–927.
- Cargnelli L and Gross MR, 1996. The temporal dimension in fish recruitment: birth date, body size, and size-

- dependent survival in a sunfish (bluegill: *Lepomis macrochirus*). Can J Fish Aquat Sci 53:360–367.
- Chakraborty R, Shaw M, and Schull WJ, 1974. Exclusion of paternity: the current state of the art. Am J Hum Genet 26:477–488.
- Chakraborty R, Meagher TR, and Smouse PE, 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118:527–536.
- Colbourne JK, Neff BD, Wright JM, and Gross MR, 1996. DNA fingerprinting of bluegill sunfish (*Lepomis macrochirus*) using (GT)_n microsatellites and its potential for assessment of mating success. Can J Fish Aquat Sci 53:342–349.
- DeWoody JA, Fletcher DE, Wilkins D, Nelson WS, and Avise JC, 1998. Molecular genetic dissection of spawning, parentage, and reproductive tactics in a population of redbreast sunfish, *Lepomis auritus*. Evolution 52:1802–1810.
- DeWoody JA, DeWoody YD, Fiumera AC, and Avise JC, 2000a. On the number of reproductives contributing to a half-sib progeny array. Genet Res Camb 75:95–105.
- DeWoody JA, Walker D, and Avise JC, 2000b. Genetic parentage in large half-sib clutches: theoretical estimates and empirical appraisals. Genetics 154:1907–1912.
- Dominey WJ, 1980. Female mimicry in male bluegill sunfish—a genetic polymorphism? Nature 284:546–548.
- Dominey WJ, 1983. Mobbing in colonially nesting fishes, especially the bluegill, *Lepomis macrochirus*. Copeia 1:1086–1088.
- Double MC, Cockburn A, Barry SC, and Smouse PE, 1997. Exclusion probabilities for single-locus paternity analysis when related males compete for matings. Mol Ecol 6:1155–1166.
- Estoup A, Gharbi K, SanCristobal M, Chevalet C, Haf-ray P, and Guyomard R, 1998. Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. Can J Fish Aquat Sci 55:715–725.
- Fu P, 2000. Sperm competition and the evolution of alternative reproductive life histories in bluegill sunfish (*Lepomis macrochirus*) (Master thesis). Toronto: University of Toronto.
- Godin J-GJ, ed, 1997. Behavioural ecology of teleost fishes. New York: Oxford University Press.
- Gross MR, 1980. Sexual selection and the evolution of reproductive strategies in sunfishes (*Lepomis*: Centrarchidae) (PhD dissertation). Salt Lake City: University of Utah.
- Gross MR, 1982. Sneakers, satellites and parentals: polymorphic mating strategies in North American sunfishes. Z Tierpsychol 60:1–26.
- Gross MR, 1984. Sunfish, salmon, and the evolution of alternative reproductive strategies and tactics in fishes. In: Fish reproduction: strategies and tactics (Potts G and Wootton R, eds). London: Academic Press; 57–75.
- Gross MR, 1991. Evolution of alternative reproductive strategies: frequency-dependent sexual selection in male bluegill sunfish. Phil Trans R Soc Lond B 332:59–66.
- Gross MR, 1996. Alternative reproductive strategies and tactics: diversity within sexes. Trends Ecol Evol 11:92–98.
- Gross MR and Charnov EL, 1980. Alternative male life histories in bluegill sunfish. Proc Natl Acad Sci USA 77:6937–6940.
- Jarne P and Lagoda JL, 1996. Microsatellites, from molecules to populations and back. Trends Ecol Evol 8:285–288.
- Lee DS, Gilbert CR, Hocutt CH, Jenkins RE, McAllister DE, and Stauffer JR Jr, 1980. Atlas of North American freshwater fishes. Raleigh: North Carolina State Museum of Natural History.
- Lukas JA and Orth DJ, 1993. Reproductive ecology of redbreast sunfish *Lepomis auritus* in a Virginia stream. J Freshwater Ecol 8:235–244.
- Marshall TC, Slate J, Kruuk LEB, and Pemberton JM, 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655.
- Maynard Smith J, 1982. Evolution and the theory of games. Cambridge: Cambridge University Press.
- 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.
- Neff BD, 2000. Genetic markers and breeding success: theoretical and empirical investigations in fish (PhD dissertation). Toronto: University of Toronto.
- Neff BD, Fu P, and Gross MR, 1999. Microsatellite evolution in sunfish (Centrarchidae). Can J Fish Aquat Sci 56:1198–1205.
- Neff BD, Fu P, and Gross MR, 2000a. Microsatellite multiplexing in fish. Trans Am Fish Soc 129:590–599.
- Neff BD, Repka J, and Gross MR, 2000b. Parentage analysis with incomplete sampling of parents and offspring. Mol Ecol 9:515–528.
- Neff BD, Repka J, and Gross MR, 2000c. Statistical confidence in parentage analysis with incomplete sampling: How many loci and offspring are needed? Mol Ecol 9:529–540.
- O'Reilly PT, Hamilton LC, McConnell SK, and Wright JM, 1996. Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Can J Fish Aquat Sci 53:2292–2298.
- Pemberton JM, Slate J, Bancroft DR, and Barrett JA, 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. Mol Ecol 4:249–252.
- Philipp DP and Gross MR, 1994. Genetic evidence of cuckoldry in bluegill *Lepomis macrochirus*. Mol Ecol 3:563–569.
- Repka J and Gross MR, 1995. The evolutionarily stable strategy under individual condition and tactic frequency. J Theor Biol 176:27–31.
- Smouse PE and Meagher TR, 1994. Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). Genetics 136:313–322.
- Taborsky M, 1997. Bourgeois and parasitic tactics: do we need collective, functional terms for alternative reproductive behaviors? Behav Ecol Sociobiol 41:361–362.
- Taborsky M, 1998. Sperm competition in fish: “bourgeois” males and parasitic spawning. Trends Ecol Evol 13:222–227.
- van Damme E, 1991. Stability and perfection of Nash equilibria. New York: Springer-Verlag.

Corresponding Editor: John C. Avise