

# Microsatellite evolution in sunfish (Centrarchidae)

Bryan D. Neff, Peng Fu, and Mart R. Gross

**Abstract:** We describe 10 microsatellite loci from bluegill (*Lepomis macrochirus*) and discuss their evolution within the Centrarchidae. All of the loci exhibit Mendelian inheritance and are unlinked. While six loci are conserved within the Centrarchidae (found also in pumpkinseed (*Lepomis gibbosus*) and largemouth bass (*Micropterus salmoides*)), four have origins outside the family and two predate it by 65–150 million years. The persistence of these loci in fish may be due to a slow rate of sequence divergence within their flanking sequences, estimated at 0.14–0.83% per million years. We examine the number of alleles, heterozygosity, range, modes, and the frequency of the most common allele and find that a two-phased model (TPM) or an infinite alleles-model (IAM) best describes the results, while a stepwise mutation model (SMM) is rejected. Therefore, population differentiation analyses utilizing these microsatellite loci should consider parameters based on the IAM (e.g.,  $F_{ST}$ ) and not the SMM.

**Résumé :** Nous décrivons 10 loci microsatellites du crapet arlequin (*Lepomis macrochirus*) et traitons de leur évolution au sein des Centrarchidae. Tous les loci obéissent à l'hérédité mendélienne et ne sont pas liés. Six loci sont conservés au sein des Centrarchidae (présents aussi chez le crapet-soleil (*Lepomis gibbosus*) et l'achigan à grande bouche (*Micropterus salmoides*)), tandis que quatre ont leur origine à l'extérieur de la famille, dont deux existaient 65 à 150 millions d'années avant l'apparition de cette dernière. La persistance de ces loci chez les poissons pourrait être attribuable à un faible taux de divergence des séquences dans leurs régions flanquantes, estimé à 0,14–0,83% par million d'années. Nous examinons le nombre d'allèles, l'hétérozygoté, la diversité des allèles, les modes et la fréquence de l'allèle le plus commun pour conclure qu'un modèle diphasé ou un modèle en nombre infini d'allèles donnent la meilleure description des résultats et qu'on doit rejeter le modèle de mutation par pas. Ainsi, les analyses de différenciation des populations utilisant ces loci microsatellites devraient considérer les paramètres fondés sur le modèle en nombre infini d'allèles (p. ex.  $F_{ST}$ ) et non sur le modèle de mutation par pas.

[Traduit par la Rédaction]

## Introduction

Microsatellite loci are powerful genetic markers with the potential to provide researchers with new insights into the behaviour, ecology, and genetic structure of fish (Avice 1994; Jarne and Lagoda 1996; O'Connell and Wright 1997). Here, we report an analysis of 10 microsatellites in the North American sunfish family Centrarchidae (eight new and two reported in Colbourne et al. 1996). We evaluate the microsatellite loci within the source species, across two species within the same genus, across two genera within the same family, and across three fish families. Our analyses infer the origins of these loci and their mode and rates of evolution as well as their expression and variability. Thus, this study provides an evolutionary perspective on microsatellites and a useful set of genetic markers.

Three prevailing models exist to explain the mode of mutation at genetic marker loci: the infinite-alleles model (IAM), the stepwise mutation model (SMM), and the two-phase model (TPM). These models relate to parameters used

to estimate population differentiation and distance and are therefore important for considerations in fisheries management and conservation. The IAM (Kimura and Crow 1964) assumes that, within a population, each mutation gives rise to a unique allele and therefore, equivalent alleles are identical by descent. The SMM (Kimura and Ohta 1978; Shriver et al. 1993; Valdes et al. 1993) assumes that a mutation changes the allele by one unit and therefore infers that alleles of the same size are more closely related but not necessarily identical by descent. The TPM (Di Rienzo et al. 1994) is analogous to the SMM with the addition of occasional multistep mutations. This latter model requires estimates of the size (i.e., number of steps) and proportion of mutations that are multistep.

It is generally considered that the mode of mutation at microsatellite loci follows a stepwise mutational process (e.g., Schlötterer and Tautz 1992). However, allelic variation at many loci appears to conform to predictions based on the IAM and not the SMM (e.g., O'Connell et al. 1997). Further, measures of population differentiation calculated from parameters based on the IAM are often consistent with alternative estimates (e.g., estimates based on mtDNA; Di Rienzo et al. 1994). Di Rienzo et al. (1994) have suggested that multistep mutations may often give rise to novel alleles and could explain the apparent fit of microsatellite data to the IAM. However, few studies have tested the TPM, and to our knowledge, no study has yet tested all three models.

The Centrarchidae diverged from other fishes during the

Received August 28, 1998. Accepted March 18, 1999.  
J14761

**B.D. Neff,<sup>1</sup> P. Fu, and M.R. Gross.<sup>2</sup>** Department of Zoology, University of Toronto, 25 Harbord Street, Toronto, ON M5S 3G5, Canada.

<sup>1,2</sup>Authors to whom all correspondence should be addressed.  
e-mail: neff@zoo.utoronto.ca; mgross@zoo.utoronto.ca

Eocene, 39–54 million years ago (Wilson and Williams 1992). The family is now composed of 32 species in nine genera (Nelson 1984; Wainwright and Lauder 1992). These species are widely distributed throughout their native North America and through human transport have colonized other regions of the world (e.g., Lelek 1987; Coad et al. 1995). They are important members of the fish fauna wherever they are found. In North America, they are highly valued in the sport fishery, being the target species in over 80% of registered sport tournaments (Schramm et al. 1991). They are also a concern to conservation biologists, both because of endangered populations in North America and because they can displace native species when introduced elsewhere (Lelek 1987). In addition, they are of great interest to evolutionary ecologists for studies of life history, mating system, and behavioural evolution (Gross 1982, 1996; Wainwright and Lauder 1992; Baylis et al. 1993; Belk 1998).

In this study of microsatellite evolution, the source species is the bluegill (*Lepomis macrochirus*). Bluegill are thought to be a recent lineage within the Centrarchidae that diverged probably 2–4 million years ago during the late Pliocene to early Pleistocene (Miller 1958; Mabee 1988; Wainwright and Lauder 1992). The pumpkinseed (*Lepomis gibbosus*), which cohabits with bluegill in the northern part of its range, is used as the test species for cross-species amplification within the genus. The largemouth bass (*Micropterus salmoides*), probably the first of the centrarchids to have diverged 39–54 million years ago during the Eocene (Wainwright and Lauder 1992; Wilson and Williams 1992), is used as the test species for cross-genera amplification and comparison between the oldest and youngest groups within the family. We examine the level of polymorphism within four largemouth bass populations and compare this with that of bluegill. To determine whether the microsatellite loci are exclusive to the Centrarchidae, we test for their presence in the families Percidae (yellow perch (*Perca flavescens*)) and Esocidae (northern pike (*Esox lucius*)), an ancient group that diverged during the late Cretaceous, 65–150 million years ago (Wilson and Williams 1992). Yellow perch and northern pike are also important recreational and commercial species. To estimate the rate of sequence divergence at the loci, we analyze sequences from homologues of two loci in bluegill, pumpkinseed, and largemouth bass. We compare the rates with those found for other species in the literature and infer the relationship between microsatellite mutation rates and conservation. Finally, we test the IAM, SMM, and TPM to determine which model best explains the variation at the microsatellite loci.

## Methods

### Fish samples

The microsatellite loci were initially isolated from a bluegill in Lake Opinicon (Ontario, Canada). For each locus, inheritance of the microsatellite alleles was tested using up to three family crosses and a minimum of 20 offspring per family (see Colbourne et al. 1996 for details of the crosses). Linkage disequilibrium was examined using genotypes from 44 bluegill of mixed ages seined from several areas of Lake Opinicon. Cross-species amplifications were tested using DNA of five pumpkinseed, five largemouth bass, two yellow perch, and two northern pike netted or angled from Lake Opinicon. Sequence analyses used a single bluegill,

pumpkinseed, and largemouth bass from Lake Opinicon. Comparison of allelic polymorphism was tested using 20 bluegill and largemouth bass taken from multiple populations to minimize population-specific biases: bluegill from Lake Opinicon ( $n = 10$ ) and Lake Fork (Texas, U.S.A.,  $n = 10$ ) and largemouth bass from Lake Opinicon ( $n = 5$ ), Everglades (Florida, U.S.A.,  $n = 5$ ), Lake Fork ( $n = 5$ ), and Lake Erie (Minnesota, U.S.A.,  $n = 5$ ). The mutational models were tested using genotypes from 169 bluegill of mixed ages seined from several areas of Lake Opinicon.

### Microsatellite protocols

We followed the protocols of Colbourne et al. (1996). Briefly, microsatellites were developed from a partial genomic library of bluegill DNA. Positive clones were sequenced and “forward” and “reverse” primers were designed for bluegill using the program PRIMER (Lincoln et al. 1991). Single-locus and multiplex polymerase chain reactions were then performed for all species using the same protocols. Two microsatellite loci were sequenced using DNA from a homozygous bluegill, pumpkinseed, and largemouth bass and a T<sub>7</sub> cycle sequencing kit (Pharmacia).

### Inheritance

Conformity to Hardy–Weinberg proportions was tested following Guo and Thompson (1992), and linkage of loci was tested using the program Genepop, version 1.2 (Raymond and Rousset 1995). Null alleles were screened by testing for an excess of homozygotes, inconsistent Mendelian inheritance patterns, and non-amplifying individuals.

### Cross-amplification

Cross-species amplifications of the bluegill microsatellite loci in pumpkinseed, largemouth bass, yellow perch, and northern pike provided intragenus, intergenera, and interfamily comparisons. Potential origins of the microsatellites were inferred from the cross-species amplifications and the known evolutionary relationships of the five species. To compare levels of variability of the microsatellites that were conserved within the Centrarchidae, four loci were amplified in a large sample ( $n = 20$ ) of both bluegill and largemouth bass. Both the heterozygosity, estimated according to Nei (1978), and the number of alleles present were compared. Finally, for both bluegill and largemouth bass, the probability of identity (i.e., the probability that two individuals would have identical genotypes) was calculated using the frequency of most common genotypes following Edwards et al. (1992).

### Mutation

We followed Rico et al. (1996) to estimate the rate of sequence divergence within the centrarchid flanking sequences. Briefly, sequences were aligned using CLUSTAL W with default settings (Thompson et al. 1994). Priming and repeat sequences were omitted. The estimate was based on the percent sequence similarity and the time since the divergence of the older *Micropterus* species (largemouth bass) and the more recent *Lepomis* species (bluegill and pumpkinseed). Estimates were made for two independent loci (Lma116 and Lma120).

We followed Di Rienzo et al. (1994) to estimate the frequency of slipped-strand mispairing. Briefly, the mutation rate ( $\mu$ ) was calculated from  $\mu = m_2 / (2N_e\sigma^2)$ , where  $m_2$  is the variance in allele size,  $N_e$  is the effective population size, and  $\sigma^2$  is the net increase in variance under mutation (e.g., under the SMM,  $\sigma^2 = 1$ ). Since we did not know the effective population size, we calculated a relative measure of the mutation rate for each locus from the relationship  $\mu \propto m_2$ .

Finally, we followed Estoup et al. (1995), Shriver et al. (1993), and Di Rienzo et al. (1994) to compare the allelic distribution of the microsatellites with those predicted by the IAM, SMM, and TPM, respectively. For the IAM, the expected number of alleles

**Table 1.** Primer and microsatellite sequences for 10 microsatellite loci isolated from bluegill.

Locus	Primer sequence <sup>a,b</sup>	Microsatellite sequence <sup>b</sup>	Cross-species amplification <sup>c</sup>	Additional analyses <sup>d</sup>
Lma21 <sup>e</sup>	F: CAGCTCAATAGTTCTGTGTCAGG R: ACTACTGCTGAAGATATTGTAG	(TC) <sub>19</sub> (AC) <sub>11</sub>	PS, LB, PE	M, P
Lma87 <sup>e</sup>	F: ATGACACAGACTCACCATGC R: CTCCTGCCCATAAATCAGAC	(AC) <sub>15</sub> A <sub>5</sub>	PS	M
Lma102	F: CTGTGAAAATGGTGTGAGCG R: AAACACAAAAGTCCACGCAC	(GT) <sub>19</sub> ATGTAT(GT) <sub>4</sub>	PS	M
Lma104	F: TCTAGCTGTAGTGCCACAGGG R: ACAGGGACGAAGATCCCAC	(GT) <sub>17</sub>	PS, LB, PE, PI	M
Lma116	F: GAATCTGGCTGCACACTGAC R: GTATTGACCCATAGCAGCTGC	(GT) <sub>25</sub>	PS, LB, PE, PI	M, P, S
Lma117 (short)	F: TCACTCTGCAACCTACAAGCC R: AGATCTACATGCCACTCATTCG	(GT) <sub>22</sub>	PS	M
Lma117 <sup>f</sup> (long)	F: CCACCAACAGCATGCAGAC R: CATGCCACTCATTGCACTG	(GT) <sub>22</sub>	PS	—
Lma120	F: TGTCCACCCAAACTTAAGCC R: TAAGCCCATTCCCAATTATCC	(GT) <sub>28</sub>	PS, LB	M, P, S
Lma121	F: CACTGGTGGGAGAGGAAGAG R: AGTGATTTTCCCCTGGCAG	(GT) <sub>25</sub>	PS, LB	M, P
Lma122 <sup>f</sup>	F: AAATAAAGTTTGGGATACCCTG R: GTCGTGAGCTATTGCTTAAATAC	(GT) <sub>30</sub>	PS	—
Lma124	F: CATTCAATTTGTCGCGACACC R: AAACACAAAAGCCACACCATG	(TG) <sub>27</sub>	PS, LB, PE	M

**Note:** Successful cross-species amplifications and loci with additional analyses are also indicated.

<sup>a</sup>F, forward; R, reverse.

<sup>b</sup>Sequences are from 5' to 3'.

<sup>c</sup>PS, Pumpkinseed; LB, largemouth bass; PE, yellow perch; PI, northern pike.

<sup>d</sup>M, mode of mutation; P, polymorphism; S, sequence (see text).

<sup>e</sup>Primer and microsatellite sequence previously reported in Colbourne et al. (1996).

<sup>f</sup>Null alleles detected (see text).

and heterozygosity was estimated according to Ewens (1972). For the SMM and TPM, simulations were performed to generate 95% confidence intervals (CI) for the number of alleles, frequency of the most common allele, range in allele size, number of modes (based on allele frequency and size), and heterozygosity. The effective population size was set at 1000 and the mutation rate was estimated according to  $\mu = m_2/2N_e\sigma^2$ . Since the SMM tests are sensitive to demographic history (Di Rienzo et al. 1994), we also examined the effects of coalescence time and population bottlenecks on the results of the simulations. The average coalescence time in generations ( $T$ ) was estimated using  $T = m_2/(\mu\sigma^2)$  (Di Rienzo et al. 1994), where, based on our data,  $m_2 \approx 11$  and assuming only single-step mutations,  $\mu \geq 10^{-2}$ , and  $\sigma^2 = 1$  (Jarne and Lagoda 1996). A modified SMM simulation that applied a bottleneck consistent with the founding time of the bluegill population in Lake Opinicon was also tested (Hudson 1990). For the TPM, we assumed that multistep mutations followed a geometric distribution (see Di Rienzo et al. 1994). We assumed that the associated change in repeat number was  $j + 1$  (thus, multistep mutations were of at least two repeats units). In accordance with Di Rienzo et al. (1994), the proportion of multistep mutations was  $\leq 20\%$  and the variance in these mutations was either 50 or 100.

All of the simulations ran for 4000 ( $4N_e$ ) generations after which 100 samples of 338 gametes were drawn and analyzed (338 is the number of gametes that were used to calculate the observed variance in allele size). Only samples that had a variance in allele size within 15% of  $m_2$ , the observed variance in allele size, were considered. The simulation was repeated 50 times, generating a total of 5000 ( $50 \times 100$ ) samples. The 95% CI was determined by extracting the 125th smallest and largest value for each summary measure (the 2.5 and 97.5 percentiles, respectively).

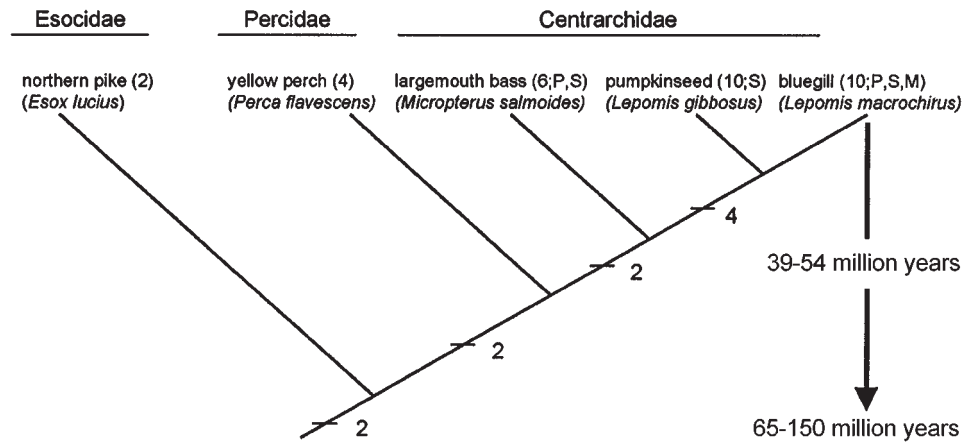
## Results

### Bluegill microsatellites

About  $1.0 \times 10^4$  colonies were screened and hybridization revealed over 200 positive clones. Sixty-five were sequenced and 85% (55/65) contained a (GT)<sub>n</sub> microsatellite. Repeats ranged from five to 75 GT's ( $25 \pm 15$  SD), and 53% (29/55) were perfect, while 47% (26/55) were imperfect repeats. Primer sequences and the repeat motif for the 10 microsatellites are presented in Table 1.

Null alleles were detected for two microsatellite loci, Lma117 (long) and Lma122. Both had an individual for which no amplified product was observed (Lma117: one in 44 individuals; Lma122: one in 44 individuals; note that these were different individuals) and an excess of homozygotes (Lma117:  $\chi^2 = 8.3$ ,  $df = 1$ ,  $p < 0.01$  ( $n = 43$ ); Lma122:  $\chi^2 = 6.3$ ,  $df = 1$ ,  $p < 0.05$  ( $n = 43$ )). For Lma117, both primers were redesigned (Lma117 (short)) and subsequent analysis revealed that the null allele consisted of a single allele with a frequency of 0.19. For Lma122, the microsatellite resided near the end of the plasmid insert and we therefore had insufficient sequence data to redesign the primers. However, by reducing the priming stringency, the null allele was amplified with weak product intensity. It consisted of at least four different alleles, all of which overlapped in length with other amplifying alleles. However, the "null" allele was less intense than the other alleles, and in some heterozygous individuals, it was undetectable through the stutter bands of

**Fig. 1.** Minimum estimates of the origin for 10 microsatellites isolated from bluegill. The phylogeny is based on the literature and the dates are based on the fossil record (see text). The number of microsatellites present in each species and a summary of the analyses (polymorphism (P), sequences (S), mode of mutation (M)) are indicated in parentheses. For example, six of the 10 microsatellites isolated in bluegill were also found in largemouth bass, and both the degree of polymorphism and sequence similarity were analyzed in the largemouth bass.



**Table 2.** Number of alleles, size range, and heterozygosity of the 10 microsatellites in bluegill (BG) and largemouth bass (LB).

Locus	Number of alleles <sup>a</sup>		Allele size range <sup>a,c</sup>		Heterozygosity <sup>a,c</sup>	
	BG	LB <sup>d</sup>	BG	LB <sup>d</sup>	BG	LB <sup>d</sup>
Lma21	5	15	158–182	124–222	0.61	0.82
Lma87	4	N	118–152	N	0.55	N
Lma102	5	N	88–102	N	0.64	N
Lma 104	2	NT	98, 100	NT	0.50	NT
Lma116	6	2	264–322	246–256	0.47	0.31
Lma117 (short)	5	N	152–176	N	0.60	N
Lma120	5	6	211–245	283–295	0.56	0.67
Lma121	4	5	176–196	174–178	0.53	0.15
Lma122	7	N	73–132	N	0.69	N
Lma 124	8	NT	152–176	NT	0.49	NT
Average	5±2	7±6	—	—	0.56±0.07	0.49±0.31

<sup>a</sup>Based on 20 individuals.  
<sup>b</sup>Allele size range is expressed as the number of nucleotides.  
<sup>c</sup>Nei (1978).  
<sup>d</sup>N, no amplification; NT, not tested (see Table 1).

**Table 3.** Flanking sequence homology of microsatellites Lma116 and Lma120 for bluegill, pumpkinseed, and largemouth bass expressed as a proportion of the nucleotides that were identical (total nucleotides given in parentheses).

	Bluegill		Pumpkinseed	
	116	120	116	120
Pumpkinseed	0.97 (176)	0.99 (76)	—	—
Largemouth bass	0.68 (157)	0.92 (103)	0.67 (179)	0.92 (75)

the normal allele. To avoid possible inaccuracies introduced by the null allele, Lma122 was only considered in the cross-species amplifications.

The remaining eight microsatellite loci and the redesigned Lma117 (short) exhibited inheritance patterns that were con-

sistent with their parents ( $\chi^2$  test,  $p > 0.20$  for each of the nine loci), conformed to the expected Hardy–Weinberg proportions (exact test,  $p > 0.40$ ), and were amplified in all individuals ( $n = 44$ ). These nine microsatellite loci averaged  $5 \pm 2$  alleles with a heterozygosity of  $0.55 \pm 0.06$ . There was no linkage between any of the loci (exact test,  $p > 0.10$  for all pairs of loci). Within the Lake Opinicon bluegill population, given the number of loci and alleles identified, the probability that two random individuals would have identical genotypes was 0.0001 ( $n = 9$  loci, mean for each locus =  $0.38 \pm 0.09$ ).

**Multiplexes**

Three sets of four microsatellites were successfully coamplified in bluegill. One set contained Lma102, Lma87, Lma21, and Lma117 (long), another contained Lma102,



**Table 4** (concluded).

Modes			Frequency (%)		
Obs	TPM	SMM	Obs	TPM	SMM
3	0–4	1–5	57	25–70	18–48
2	0–3	2–6	49	36–86	14–32
4	1–4	2–6	41	21–68	13–36
1	1–2	0–1	53	52–93	50–78
3	1–5	3–8	64	23–70	9–24
3	0–4	2–6	45	31–91	13–30
4	1–4	1–5	53	20–59	17–41
3	1–4	1–5	66	22–68	17–42
3	1–5	1–5	69	22–70	16–41
	100	100		100	11

The polymorphism of the microsatellite loci was determined for bluegill and largemouth bass (Table 2). It was not significantly different for the four microsatellites examined in both species. At these loci, the average number of alleles in bluegill and largemouth bass was  $5.0 \pm 0.8$  and  $7.0 \pm 5.6$  (paired *t* test,  $t = 0.69$ ,  $df = 3$ ,  $p = 0.54$ ), respectively, and their expected heterozygosity was  $0.54 \pm 0.06$  and  $0.49 \pm 0.31$  (paired *t* test,  $t = 0.41$ ,  $df = 3$ ,  $p = 0.71$ ), respectively. The probability that two random individuals had identical genotypes in the largemouth bass was 0.015 ( $n = 4$  loci, mean for each locus =  $0.45 \pm 0.30$ ).

### Mutation

Large portions of the Lma116 and Lma120 sequences were similar in the bluegill, pumpkinseed, and largemouth bass (Fig. 2; Table 3). Alignments revealed significant homology, although substitution, deletion, and insertion mutations were observed. Lma120 had higher overall homology than Lma116 in both the repeat region and the flanking regions (average homology: Lma116, 0.77; Lma120, 0.94). Lma116 showed substantial degeneration of the repeat. For both loci, the repeat region was shorter in the nonsource species than that of the source species (Lma116: bluegill, 25 repeats; pumpkinseed, 11; largemouth bass, 12; Lma120: bluegill, 28; both others, 13). Based on the sequence similarity of all three centrarchid species, and assuming a divergence time of 39–54 million years, the estimated rate of sequence divergence at Lma116 was  $0.83 \pm 0.02$  to  $0.59 \pm 0.01\%$  per million years and at Lma120 was  $0.20 \pm 0.00$  to  $0.14 \pm 0.00\%$  per million years.

The variance in repeat number at the nine microsatellite loci ranged from 0.25 at Lma104 to 31.7 at Lma116 (Table 4). The large variance observed at Lma116 was attributed to two large gaps in allele sizes, of seven and 12 repeats, resulting in three “clusters” of alleles at this locus. The average variance was  $11 \pm 9$  repeats. Since the rate of slipped-strand mispairing is directly proportional to the variance in repeat number, it showed over 100-fold variation ( $31.7/0.25$ ) among the microsatellite loci.

The results from the IAM, SMM, and TPM analyses are summarized in Table 4. The nine microsatellite loci fit the

TPM consistently, as it was possible to select values for the parameters  $p$  and  $\sigma^2$  such that the microsatellite characters fit the model (i.e., fall within the 95% CI). For example, at  $p = 0.05$  and  $\sigma^2 = 50$ , 56% (5/9) of the loci fell within the expected number of alleles, 67% (6/9) within the heterozygosity, 89% (8/9) within the range, 100% (9/9) within the number of modes, and 78% (7/9) within the frequency of the most common allele. All together, 44% (4/9) of the loci conformed to the TPM with these parameters at all five characters. Three of the remaining five loci conformed to the TPM at each of the five characters for  $p = 0.20$  and  $\sigma^2 = 100$ . The last two loci fit the TPM at values of  $p < 0.05$ . In contrast, the observed microsatellite data deviated significantly from that expected under the SMM. Only one locus conformed to the SMM expectation at all of the characters. All nine microsatellite loci also conformed to the IAM for heterozygosity and 78% (7/9) conformed to the IAM for the number of alleles.

## Discussion

### Microsatellites

This study provides primers for 10 microsatellite loci that should be useful for centrarchid research including applications such as population differentiation and the study of life history, mating system, and behavioural evolution. The microsatellites are polymorphic, averaging over five alleles and a heterozygosity of 0.56 per locus. The low probability of identity associated with these loci should also make them useful for evolutionary and ecological studies requiring, for example, the tracking of individuals and parentage analyses.

Initially, two of the 10 microsatellite loci had a null allele. We were only able to inconsistently visualize the “null” alleles by reducing the primer annealing stringency. For one locus, we also redesigned primers and were successful at consistently amplifying its null allele. Therefore, redesigning primers may be a better approach to consistently detect a null allele (also see Paetkau and Strobeck 1995).

Some of the microsatellite loci are also found in the pumpkinseed, largemouth bass, yellow perch, and northern pike. Those that were conserved in largemouth bass were equally polymorphic as compared with bluegill, the microsatellite source species. It is likely that these loci will have similar levels of polymorphism in all of the species within the sunfish family, since largemouth bass are most distantly related to bluegill (e.g., see Primmer et al. 1996 for a discussion). The multiplexes that we have developed for bluegill should increase the efficiency of generating genotype data using the microsatellites and may also facilitate their use with other centrarchid species.

### Evolution

Microsatellite loci generally have ancient origins and show considerable evolutionary conservation. For example, between freshwater and marine species of turtles, microsatellite homologues have shown at least 300 million years of conservation (FitzSimmons et al. 1995). In fish, homologues have been found between species from agnathans and gnathostomes, which diverged about 470 million years ago (Rico et al. 1996). Consistent with the long persistence of microsatellite loci, we found that two of the loci isolated

from bluegill were present in northern pike, indicating origins predating 65–150 million years. However, several of the loci have diverged sufficiently to be undetectable even in closely related species of the same genus (e.g., Lma87, Lma102, Lma117, and Lma122 were common only to bluegill and pumpkinseed). This suggests that not all microsatellite loci from poikilotherms show sufficient conservation of priming sites to be detectable over long periods of time.

The rate of sequence divergence of nuclear DNA is thought to be lower in poikilotherms than in homeotherms (Rico et al. 1996 and references within). For example, Rico et al. (1996) estimated the rates of sequence divergence for nuclear microsatellite flanking sequences in fish to be between 0.02 and 0.05% per million years, while in mammals, Pépin et al. (1995) estimated a rate of about 0.93% per million years. Our estimate of the rates of sequence divergence are 0.14–0.83% per million years, intermediate to the estimates in mammals and fish. Therefore, it is not clear whether there will be a consistent difference between poikilotherms and homeotherms. Further, our data suggest that this rate can be quite variable within a species. This may suggest that the rate of sequence divergence at microsatellite flanking sequences is dependent on the locus itself.

Our tests of the mode of mutation for microsatellites support the TPM and IAM while rejecting the SMM. The failure of the SMM to explain the observed variations in our data could not be attributed to a recent bottleneck. Bluegill entered Lake Opinicon only about 170 years ago (Curran et al. 1947). If this population was founded by only a few individuals (i.e., bottlenecked) followed by rapid population growth, it is expected that the time of coalescence would be close to the time of the bottleneck (Di Rienzo et al. 1994 and references within). However, our estimate of  $T$  (the average coalescence time), based on the SMM, is  $\geq 1100$  generations. This is considerably longer than the age of the Lake Opinicon population, suggesting that the population has not experienced a recent bottleneck. Furthermore, the observed data did not consistently fall within the predicted ranges based on the SMM simulations even when a population bottleneck was incorporated. For example, when assuming that the population was founded by 0.5% of the source population, only 22% (2/9) of the microsatellites conformed to all five microsatellite characteristics. Although we cannot exclude other historical effects on demography, it does not appear that a recent bottleneck can explain the observed rejection of the SMM for these microsatellites.

Although it is generally considered that microsatellite variation is generated via a stepwise mutational process (e.g., Schlötterer and Tautz 1992), our results indicate that this process may not involve only single repeats. Instead, the conformity to the TPM suggests that these microsatellite loci follow a stepwise mutational process with the occasional large mutation of multiple repeats. These large mutations are more likely to generate alleles of novel sizes than single-step mutations and may explain the conformity of such microsatellites to the IAM, since it assumes that mutations give rise to novel alleles (also see Di Rienzo et al. 1994). However, the TPM and IAM are not equivalent models. The IAM assumes that alleles of equivalent size are identical by descent, while no relationship can be inferred among alleles of different sizes. The TPM does not assume that alleles of

equivalent size are identical by descent but that alleles of similar sizes are more closely related.

It is necessary to determine which mutation model best describes microsatellite allelic variation in order to select the appropriate test statistics for estimating population differentiation and distance. For example, Slatkin (1995) showed that unlike  $R_{ST}$ ,  $F_{ST}$  tends to underestimate the true level of population differentiation when using microsatellites that conform to the SMM. As statistics based on the TPM are not yet available, those based on the IAM may be the most appropriate for the presented microsatellites that conform to both the IAM and TPM. However, since the IAM does not necessarily reflect the stepwise mutational process of microsatellites, the development of test statistics based on the TPM may provide more accurate estimations of population differentiation and structure. To our knowledge, this study represents that first test of all three mutation models showing the conformity of empirical data to both the TPM and IAM.

## Acknowledgments

John Colbourne, Junjian Chen, and Paul Hebert are thanked for their advice and assistance in forming the DNA library. Lucy Dueck of the Texas Freshwater Fisheries Center and Douglas W. Nelson of the University of Michigan Museum of Zoology are thanked for supplying tissue samples. Ben Kim assisted with primer development. This work was supported by an NSERC research grant to M.R. Gross and an NSERC graduate fellowship to B.D. Neff.

## References

- Avise, J.C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York.
- Baylis, J.R., Wiegmann, D.D., and Hoff, M.H. 1993. Alternative life histories of smallmouth bass. *Trans. Am. Fish. Soc.* **122**: 500–510.
- Belk, M.C. 1998. Predator-induced delayed maturity in bluegill sunfish (*Lepomis macrochirus*): variation among populations. *Oecologia*, **113**: 203–209.
- Coad, B.W., Waszczuk, H., and Labignan, I. 1995. Encyclopedia of Canadian fishes. Canadian Museum of Nature and Canadian Sport Fishing Productions Inc., Ottawa, Ont.
- Colbourne, J.K., Neff, B.D., Wright, J.M., and Gross, M.R. 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.
- Curran, H.W., Bardach, J., Bowman, R.I., and Lawler, H.G. 1947. A biological survey of Lake Opinicon. Progress report. Hanson and Edgar, Kingston, Ont.
- Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M., and Freimer, N.B. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 3166–3170.
- Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T., and Chakraborty, R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, **12**: 241–253.
- Estoup, A., Tailliez, C., Cornuet, J.-M., and Solignac, M. 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol. Biol. Evol.* **12**: 1074–1084.

- Ewens, W.J. 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**: 87–112.
- FitzSimmons, N.N., Moritz, C., and Moore, S.S. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.* **12**: 432–440.
- Gross, M.R. 1982. Sneakers, satellites and parentals: polymorphic mating strategies in North American sunfishes. *Z. Tierpsychol.* **60**: 1–26.
- Gross, M.R. 1996. Alternative reproductive strategies and tactics: diversity within sexes. *Trends Ecol. Evol.* **11**: 92–98.
- Guo, S.W., and Thompson, E.A. 1992. Performing the exact test of Hardy–Weinberg proportions for multiple alleles. *Biometrics*, **48**: 361–372.
- Hudson, R.R. 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**: 1–44.
- Jarne, P., and Lagoda, J.L. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* **8**: 285–288.
- Kimura, M., and Crow, J.F. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, **49**: 725–738.
- Kimura, M., and Ohta, T. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Natl. Acad. Sci. U.S.A.* **75**: 2868–2872.
- Lelek, A. 1987. The freshwater fishes of Europe. Vol. 9. Threatened fishes of Europe. AULA-Verlag GmbH, Wiesbaden, Germany.
- Lincoln, S.E., Daly, M.J., and Lander, E.S. 1991. PRIMER: a computer program for automatically selecting PCR primers. Version 0.5. Whitehead Institute for Biochemical Research, Cambridge, MA 02142.
- Mabee, P.M. 1988. Supraneural and predorsal bones in fishes: development and homologies. *Copeia*, **1988**: 827–838.
- Miller, R.R. 1958. Origin and affinities of the freshwater fish fauna of western North America. *In Zoogeography. Edited by C.L. Hubbs. Am. Assoc. Adv. Sci. Publ. No. 51.* pp. 187–222.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**: 583–590.
- Nelson, J.S. 1984. *Fishes of the world*. 2nd ed. John Wiley & Sons, New York.
- O'Connell, M., and Wright, J.M. 1997. Microsatellite DNA in fishes. *Rev. Fish Biol. Fish.* **7**: 331–363.
- O'Connell, M., Danzmann, R.G., Cornuet, J.-M., Wright, J.M., and Ferguson, M.M. 1997. Differentiation of rainbow trout (*Oncorhynchus mykiss*) populations in Lake Ontario and the evaluation of the stepwise mutation and infinite allele mutation models using microsatellite variability. *Can. J. Fish. Aquat. Sci.* **54**: 1391–1399.
- Paetkau, D., and Strobeck, C. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol. Ecol.* **4**: 519–520.
- Pépin, L., Amigues, Y., Léplinge, A., Berthier, J., Bensaid, A., and Vaiman, D. 1995. Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. *Heredity*, **74**: 53–61.
- Primmer, C.R., Moller, A.P., and Ellegren, H. 1996. A wide-range survey of cross-species microsatellite amplification in birds. *Mol. Ecol.* **5**: 365–378.
- Raymond, M., and Rousett, F. 1995. Genpop (version-1.2). Population genetics software for exact tests and ecumenicism. *J. Hered.* **86**: 248–249.
- Rico, C., Rico, I., and Hewitt, G. 1996. 470 million years of conservation of microsatellite loci among fish species. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **263**: 549–557.
- Schlötterer, C., and Tautz, D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**: 211–215.
- Schramm, H.L., Jr., Armstrong, M.L., Funicelli, N.A., Green, D.M., Lee, D.P., Manns, R.E., Jr., Taubert, B.D., and Waters, S.J. 1991. The status of competitive sport fishing in North America. *Fisheries (Bethesda)*, **16**: 4–12.
- Shriver, M.D., Jin, L., Chakraborty, R., and Boerwinkle, E. 1993. VNTR allele frequency distribution under the stepwise mutation model. *Genetics*, **134**: 983–993.
- Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**: 457–462.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Valdes, A.M., Slatkin, M., and Freimer, N.B. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, **133**: 737–749.
- Wainwright, P.C., and Lauder, G.V. 1992. The evolution of feeding biology in sunfishes (Centrarchidae). *In Systematics, historical ecology, and North American freshwater fishes. Edited by R.L. Mayden.* Stanford University Press, Stanford, Calif. pp. 472–491.
- Wilson, M.V.H., and Williams, R.G. 1992. Phylogenetic, biogeographic, and ecological significance of early fossil records of North American freshwater teleostean fishes. *In Systematics, historical ecology, and North American freshwater fishes. Edited by R.L. Mayden.* Stanford University Press, Stanford, Calif. pp. 224–246.