# POPULATION STRUCTURE AND EFFECTIVE SIZE IN CRITICALLY ENDANGERED CAPE FEAR SHINERS *NOTROPIS MEKISTOCHOLAS*

JOHN R. GOLD[1,*], ERIC SAILLANT[1], CHRISTOPHER P. BURRIDGE[1], AIMEE BLANCHARD[1], AND JOHN C. PATTON[1]

ABSTRACT - Allelic variation at hypervariable, nuclear-encoded loci and mitochondrial (mt)DNA was studied among three geographic samples (40 individuals) of the critically endangered Cape Fear shiner, *Notropis mekistocholas*. Genetic variation, as measured by allelic richness and gene (microsatellite) or nucleon (mtDNA) diversity, was similar to that in other fish species. Homogeneity tests of allele and genotype distributions and analysis of molecular variance (AMOVA) at nuclear-encoded loci revealed significant genetic heterogeneity among localities. No differences in mtDNA allele (haplotype) frequencies were detected. The ratio of the number of microsatellite alleles to the range in allele size suggested that significant reductions in effective size have occurred at two of the three localities. Long-term (inbreeding) effective population size differed among the samples and ranged from ~1,300 to ~3,000. Collectively, these results indicate that (i) Cape Fear shiners at these localities are not genetically impoverished, (ii) separate populations of Cape Fear shiners may exist in the Cape Fear drainage, (iii) recent reduction in effective size may have occurred in two of the three localities, and (iv) ancestral populations of Cape Fear shiners may have been of sufficient effective size to offset extinction due to genetic factors.

## INTRODUCTION

The Cape Fear shiner, *Notropis mekistocholas* Snelson (1971), is a small, herbivorous cyprinid fish (minnow) known only from the Cape Fear River drainage in the east-central Piedmont region of North Carolina. The species currently is designated as critically endangered (Hilton-Taylor 2000), and is thought to exist in only five localities within the drainage (Alderman 1995). Cape Fear shiners have not been observed at two of these localities (both in the Haw River) in recent years (D. Rabon, U.S. Fish and Wildlife Service, pers. comm.), and very recent attempts to obtain Cape Fear shiners at a third locality, the Deep River above the dam at High Falls in Moore County, have been largely unsuccessful (A. Howard, NC State University, pers. comm.). While it may be that Cape Fear shiners have always existed in low numbers (Snelson 1971), Pottern and Huish (1986, 1987) reported declines in abundance and range for the species that were not evident

[1]Center for Biosystematics and Biodiversity, Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX 77843-2258. [*]Corresponding author - goldfish@tamu.edu.

for other sympatric taxa. Recently, Howard et al. (2002) found that the physical microhabitat for Cape Fear shiners was in short supply because of loss of habitat and reduced and variable water quality.

In this paper, we continue our genetic studies of extant Cape Fear shiners. Previously, Burridge and Gold (2003) surveyed allelic variation at eleven, hypervariable nuclear-encoded loci among samples from three of the known localities. Genetic variation, as measured by allele number and observed heterozygosity, was relatively high and genetic divergence among samples was minimal. Tests of heterozygosity excess (which presumably reflect reduction in effective population size) were non-significant. Collectively, these results suggested that Cape Fear shiners at the localities sampled represented the same population and were not compromised genetically. However, statistical power of these analyses was impacted both by small sample size and by the number of loci employed. Because analyzing additional (independent) genetic markers generally has greater power benefit than sampling more genomes (Cornuet and Luikart 1996), and given the difficulty in obtaining Cape Fear shiners, we chose to increase the number of markers for the specimens currently on hand. Eleven additional nuclear-encoded (hypervariable) loci (microsatellites) were developed for this purpose. We also incorporated analysis of sequences of mitochondrial (mt)DNA, bringing the total number of genetic markers assayed to 23. Our objectives, as before, were to (i) quantify genetic diversity as a means to assess whether small population effects might be occurring, (ii) determine whether the geographic samples of *N. mekistocholas* differed genetically from one another, and (iii) assess whether reductions in genetic effective size had occurred recently at the sample localities. We also used a coalescent-based approach to estimate inbreeding or long-term effective size ($N_{el}$) in each sample.

## MATERIALS AND METHODS

Geographic details of sample localities and number of specimens were given in Burridge and Gold (2003). A map of the area is presented in Figure 1. Samples included 10 individuals from the Deep River above the dam at High Falls and adjacent to the bridge on State Road 1456 (SR1456), 15 individuals downstream of the dam at High Falls (High Falls), and 15 individuals at the confluence of the Deep and Rocky rivers (Confluence). Recent attempts to sample Cape Fear shiners at the SR1456 locality have been unsuccessful. Details regarding isolation of DNA, construction of the genomic library, and design of polymerase-chain-reaction (PCR) primers for hypervariable, nuclear-encoded sequences, also were given in Burridge and Gold (2003). For this study, eleven new PCR primer pairs were designed; all amplified tandemly

repeated di-nucleotide (microsatellite) sequences (Appendix 1). PCR reaction and electrophoretic conditions followed those in Heist and Gold (2000) except that 35 cycles and 60 sec extensions were employed.

For analysis of mtDNA, a fragment of ~2,350 base-pairs (bp) that included parts of the ND-5 and ND-6 genes was amplified (60 °C annealing temperature, using primers L12301 [5'-AATAGCTCATCCRTTGGTCTTAGG-3'] and H14705 [5'-AACAACGGTGGTTTTTCAARTC-3']) and sequenced from five individuals in order to identify variable sites within the two genes. The external primers were used as sequencing primers. Primer pairs 306 (306L 5'-AACTGAGCAGATACCTATGTAAAAACT-3'; 306H 5'-AAGAGTAATTATGGCTACCAGGAATA-3') and 408 (408H 5'-TTGCTGTTGCGTCCAATCCTAC-3'; 408L 5'-ACAACAGTACCCAAGCACAAATAACTAA-3') were then designed to bracket identified polymorphic sites in the two genes. The 306 (ND-5) and 408 (ND-6) primer pairs were used as both amplification and sequencing primers. PCR reactions (50 µl) contained 5-50 ng DNA in 10 mM Tris-HCl, pH 8.7, 50 mM KCI, 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 2 µM of each of the two primers, and 1.25 Units of AmpliTaq™ DNA polymerase (Perkin Elmer, Norwalk, CT). Reactions were heated to 95 °C for 3 min followed by 35 cycles of amplification. Each cycle consisted of 45 s at 95 °C, 30 s at 50 °C and 1 min at 70 °C. PCR reactions were cleaned with sephadex columns (QIAquick PCR purification kit, Qiagen, Inc.) and sequenced with Big Dye (Applied Biosystems) following protocols suggested by the supplier. Sequences were aligned and analyzed using Sequencher® software (Genecodes Corp.).

Genetic variability for both nuclear-encoded loci and mtDNA was measured as number of alleles, allelic richness (a measure of the number of alleles independent of sample size), and gene (microsatellites) and nucleon (mtDNA) diversity. Gene and nucleon diversity were estimated

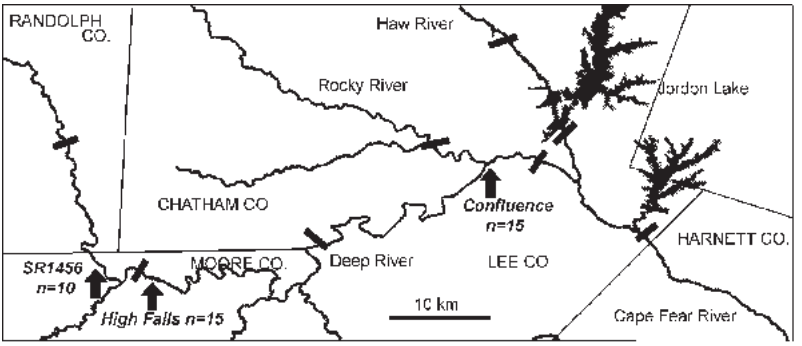


Figure 1. Collection sites (arrows) and sample sizes of *Notropis mekistocholas* from the Cape Fear drainage, NC. Bars indicate dams that may be barriers to upstream movement of *N. mekistocholas*. County boundaries are indicated.

after (Nei 1987). Homogeneity of allelic richness and gene diversity between pairs of samples was examined via Wilcoxon signed-ranks tests. Deficiency/excess of heterozygotes ($F_{IS}$) at each nuclear-encoded locus within each sample was estimated via the *f* statistic of Weir and Cockerham (1984). Estimates of allelic richness, gene and nucleon diversity, and $F_{IS}$ were obtained using F-STAT version 2.9.3.2 (www.unil.ch/izea/softwares/fstat.html). Tests for conformance of genotype proportions (nuclear-encoded loci) to Hardy-Weinberg (HW) equilibrium expectations employed an unbiased estimate of Fisher's exact-test statistic calculated by a Markov-chain procedure (1,000 dememorizations, 100 batches, and 1,000 iterations per batch). Genotypic disequilibrium between pairs of nuclear-encoded loci was tested via exact tests. Tests of HW and genotypic disequilibrium were carried out using GENEPOP 3.3 (Raymond and Rousset 1995). For mtDNA, nucleotide diversity, the number of nucleotide differences per site between two randomly chosen sequences, was estimated after Nei (1987), as implemented in ARLEQUIN 2.000 (Schneider et al. 2000).

Homogeneity of allele and genotype distributions among samples and in pairwise comparisons was tested via exact tests (as described above and using GENEPOP); homogeneity of allele distributions also was tested by analysis of molecular variance or AMOVA (Excoffier et al. 1992), as implemented in ARLEQUIN. AMOVA was conducted on each locus (including mtDNA) separately and on all nuclear-encoded loci combined; distances between alleles were calculated as number of different alleles (IAM or infinite allele model, nuclear-encoded loci and mtDNA), or sum-of-squared size difference (SMM or stepwise mutation model, 17 microsatellites only; see below). Significance of fixation index ($\Phi_{ST}$) values obtained from AMOVA was estimated from 10,100 permutations of genotypes.

For nuclear-encoded loci, the "*M*" test (Garza and Williamson 2001) was used to assess whether recent reductions in effective size had occurred in each sample. Briefly, *M* is the mean ratio of the number of alleles to the range in allele size; under the SMM model, *M* is expected to decrease when effective size of a population is reduced (Garza and Williamson 2001). We used software (M_P_Val.exe and Critical_M.exe) available at www.pfeg.noaa.gov/tib/staff/carlos_garza/carlos software.html to estimate and test significance of *M*, and parameterized the generalized SMM model as recommended in Garza and Williamson (2001), i.e., $p_s$ (proportion of one-step mutations) = 90% and $\Delta_g$ (average size of non one-step mutations) = 3.5. Values of $j$ ($4N_e\mu$, where $N_e$ = inbreeding effective population size and $\mu$ = mutation rate) were obtained empirically during estimation of inbreeding effective size (see below). Values of $\Theta$ employed were 0.66 (Confluence), 0.48 (High Falls), and 1.05 (SR1456). Only those

microsatellites (17 total) whose alleles fit assumptions of the SMM were used in the "*M*" test.

Maximum-likelihood estimates of long-term (inbreeding) effective population size ($N_{el}$) for each sample locality were generated from microsatellite data, using the program MIGRATE (Beerli 2002, Beerli and Felsenstein 2001). Estimation was based on expansion of the coalescent theory (Kingman 1982), including migration (Beerli and Felsenstein 2001), where estimates of $N_{el}$ are integrated over the time to common ancestry of all alleles in a population (Avise 2000). The parameter ($\Theta$) estimated by MIGRATE is equal to $4N_{el}\mu$, where $\mu$ is the rate of substitution per generation at a genetic locus or loci. Gene genealogies, based on observed genotypes and the estimated value of $\Theta$, are simulated, and a Monte Carlo Markov Chain (MCMC) sampling approach is used to estimate the likelihood surface of $\Theta$. Computations were performed on a SGI origin 3800 computer (Silicon Graphics Inc.) by using a parallel-processing version of MIGRATE that was run simultaneously on four processors. Each nuclear-encoded locus was analyzed separately using a Brownian approximation of the stepwise mutation model. The MCMC search employed 10 short chains (sampling 500 gene trees) and two long chains (5000 gene trees); the first $10^4$ steps were ignored in each run in order to ensure parameter stability. The analysis was repeated twice for each locus, and resulting integrated estimates of $\Theta$ were used to initialize a more thorough analysis (four replicates of each locus, using four long chains). Following parallel analysis of each locus, likelihood surfaces were integrated across loci, using MIGRATE. Estimates of $N_{el}$ and their 95% confidence intervals (CI) were derived from $\Theta$ values and their 95% likelihood-profile CI (Beerli 2002), respectively, by assuming an average microsatellite mutation rate of $10^{-4}$ (Jarne and Lagoda 1996).

## RESULTS

### Genetic variation:

Summary statistics, including sample size, number of alleles, allelic richness, gene diversity (expected heterozygosity), results of tests of genotypic conformance to Hardy-Weinberg expectations, and estimates of $F_{IS}$, for each of the 11 "new" microsatellites assayed in each of the three samples are presented in Table 1. Allele frequencies at each microsatellite in each sample may be obtained from the corresponding author. Summary data for 11 other nuclear-encoded loci (10 microsatellites and one anonymous locus) for the same samples may be found in Burridge and Gold (2003). All 11 "new" loci were polymorphic in each sample; allele number ranged from 5 to 11, with private alleles [28 total] occurring at each locality. Gene diversity (expected heterozygosity) in all samples was relatively high, with the lowest values being 0.648 (Confluence) and 0.659 (High Falls) at microsatellite *Nme*

15F2.174. No significant differences in allele richness, a measure of allele diversity independent of sample size (El Mousadik and Petit 1996), or gene diversity were found in pairwise comparisons of samples ($0.689 < P < 0.884$ and $0.205 < P < 0.758$, respectively).

Only two of 33 simultaneous tests for genotypic conformance to HW expectations differed significantly from zero following Bonferroni correction (Table 1). Both "significant" tests involved microsatellite *Nme* 12D8.125 (in High Falls and SR1456). Similarly, only $F_{IS}$ values at *Nme*

Table 1. Summary data for eleven nuclear-encoded microsatellites in three geographic samples of *Notropis mekistocholas*: sample size (N); number of alleles (A), with number of private alleles in brackets; allelic richness ($A_R$); gene diversity or expected heterozygosity ($H_E$); probability of deviation from Hardy-Weinberg equilibrium ($P_{HW}$); and deficiency/excess of heterozygotes ($F_{IS}$).

| | Sample | | | | | | |
|---|---|---|---|---|---|---|---|
| Microsatellite | locality | N | A | $A_R$ | $H_E$ | $P_{HW}$ | $F_{IS}$ |
| *Nme* 5B10.211 | Confluence | 15 | 10 [4] | 8.60 | 0.888 | 0.945 | 0.024 |
| | High Falls | 15 | 7 [1] | 6.38 | 0.850 | 0.596 | –0.020 |
| | SR1456 | 10 | 6 [1] | 5.89 | 0.596 | 0.264 | 0.285 |
| *Nme* 27F8.230 | Confluence | 15 | 7 [0] | 6.18 | 0.743 | 0.845 | –0.256 |
| | High Falls | 15 | 8 [0] | 6.88 | 0.845 | 0.477 | 0.054 |
| | SR1456 | 10 | 6 [0] | 5.89 | 0.850 | 0.448 | 0.176 |
| *Nme* 12D8.125 | Confluence | 15 | 8 [0] | 7.34 | 0.886 | 0.715 | 0.172 |
| | High Falls | 15 | 8 [0] | 6.82 | 0.852 | 0.000* | 0.844 |
| | SR1456 | 10 | 7 [0] | 6.70 | 0.850 | 0.000* | 0.529 |
| *Nme* 15F2.174 | Confluence | 15 | 6 [1] | 4.90 | 0.648 | 0.607 | –0.338 |
| | High Falls | 15 | 5 [0] | 4.67 | 0.659 | 0.114 | 0.179 |
| | SR1456 | 10 | 6 [0] | 5.80 | 0.811 | 0.734 | 0.014 |
| *Nme* 30F12.208 | Confluence | 15 | 8 [2] | 6.25 | 0.790 | 0.333 | 0.072 |
| | High Falls | 14 | 8 [1] | 7.04 | 0.871 | 0.011 | 0.344 |
| | SR1456 | 9 | 6 [0] | 6.00 | 0.799 | 0.042 | 0.304 |
| *Nme* 33B6.125 | Confluence | 15 | 5 [0] | 4.56 | 0.695 | 0.833 | –0.055 |
| | High Falls | 15 | 6 [1] | 5.24 | 0.700 | 0.787 | 0.048 |
| | SR1456 | 10 | 5 [0] | 4.90 | 0.778 | 0.479 | –0.286 |
| *Nme* 2B10.166 | Confluence | 14 | 6 [0] | 4.93 | 0.742 | 0.489 | 0.037 |
| | High Falls | 12 | 5 [0] | 4.49 | 0.655 | 0.731 | 0.110 |
| | SR1456 | 9 | 5 [0] | 5.00 | 0.715 | 0.860 | 0.068 |
| *Nme* 2B10.232 | Confluence | 15 | 9 [3] | 7.29 | 0.812 | 0.184 | 0.015 |
| | High Falls | 15 | 9 [1] | 7.58 | 0.852 | 0.518 | –0.095 |
| | SR1456 | 10 | 8 [2] | 7.69 | 0.867 | 0.227 | –0.038 |
| *Nme* 5G4.175 | Confluence | 15 | 9 [2] | 7.04 | 0.793 | 0.090 | 0.159 |
| | High Falls | 15 | 7 [1] | 5.63 | 0.752 | 0.449 | –0.063 |
| | SR1456 | 10 | 5 [1] | 4.79 | 0.689 | 1.000 | –0.161 |
| *Nme* 6A7.93 | Confluence | 15 | 11 [4] | 8.62 | 0.874 | 0.596 | –0.068 |
| | High Falls | 15 | 8 [1] | 7.07 | 0.860 | 0.462 | 0.147 |
| | SR1456 | 10 | 6 [0] | 5.69 | 0.661 | 0.891 | –0.361 |
| *Nme* 2D5.117 | Confluence | 15 | 8 [2] | 6.64 | 0.836 | 0.927 | –0.117 |
| | High Falls | 15 | 9 [0] | 7.42 | 0.843 | 0.944 | –0.028 |
| | SR1456 | 10 | 8 [0] | 7.69 | 0.889 | 0.114 | 0.213 |

*$P < 0.05$ after Bonferroni correction.

12D8.125 in High Falls ($F_{IS}$ = 0.844, P = 0.001) and SR1456 ($F_{IS}$ = 0.529, P = 0.002) differed significantly from zero. In both cases, $F_{IS}$ values were positive, indicating a deficit of heterozygotes and the possible occurrence of null alleles at this microsatellite. Because null alleles, when present, may result in erroneous inferences regarding allele and genotype distributions, microsatellite *Nme* 12D8.125 was omitted from all further analyses. Tests of genotypic disequilibrium between pairs of loci (*Nme* 12D8.125 excluded) within each sample were nonsignificant following Bonferroni correction.

A total of 625 base pairs (bp) of protein-coding mtDNA sequences was acquired from 38 specimens. The sequences included a 258 bp fragment from the ND-5 gene and a 367 bp fragment from the ND-6 gene. The ND-5 fragment corresponds to mtDNA bases 12,971–13,230 of *Cyprinus carpio* (GenBank Accession NC001606) and includes a triplet insertion at bases 12,989–12,991; the ND-6 fragment corresponds to mtDNA bases 14,777–15,143 of *C. carpio* and includes a triplet insertion at bases 14,911–14,913 and triplet deletions at 14,930–14,932 and 15,020–15,022.

A total of thirteen different mtDNA alleles (haplotypes) were identified. Summary statistics, including the distribution of mtDNA haplotypes among samples, number of polymorphic sites, nucleon and nucleotide diversity, and mean pairwise (sequence) differences between haplotypes are given in Table 2. Haplotypes 1, 2, and 7 were the most

Table 2. Summary data for mitochondrial (mt)DNA variation in three geographic samples of *Notropis mekistocholas*. Data are from 636 base pairs of mtDNA protein-coding sequence. GenBank accession numbers are for ND-5 (left) and ND-6 (right) sequences.

| | Sample locality | | | |
|---|---|---|---|---|
| Haplotype | Confluence (n = 15) | High Falls (n = 14) | SR1456 (n = 9) | Total |
| 1 (AY396545, AY396555) | 4 | 3 | 3 | 10 |
| 2 (AY396546, AY396555) | 3 | 2 | 2 | 7 |
| 3 (AY396547, AY396556) | 2 | 0 | 0 | 2 |
| 4 (AY396548, AY396556 | 1 | 0 | 0 | 1 |
| 5 (AY396545, AY396557) | 1 | 0 | 0 | 1 |
| 6 (AY396549, AY396558) | 1 | 0 | 0 | 1 |
| 7 (AY396550, AY396559) | 1 | 4 | 2 | 7 |
| 8 (AY396551, AY396560) | 1 | 1 | 0 | 2 |
| 9 (AY396552, AY396556) | 1 | 2 | 0 | 3 |
| 10 (AY396545, AY396561) | 0 | 1 | 0 | 1 |
| 11 (AY396553, AY396556) | 0 | 1 | 0 | 1 |
| 12 (AY396554, AY369560) | 0 | 0 | 1 | 1 |
| 13 (AY396545, AY396562) | 0 | 0 | 1 | 1 |
| Number haplotypes | 9 | 7 | 5 | 13 |
| Polymorphic sites | 13 | 12 | 8 | 17 |
| Nucleon diversity | 0.91 ± 0.05 | 0.88 ± 0.06 | 0.86 ± 0.09 | |
| Nucleotide diversity | $4.3 \times 10^{-3}$ | $4.5 \times 10^{-3}$ | $3.6 \times 10^{-3}$ | |
| Average pairwise difference | 2.72 | 2.85 | 2.28 | |

common, occurring (overall) in frequencies of 0.263, 0.184, and 0.184, respectively. There were seven private haplotypes, three in Confluence, and two each in High Falls and SR1456. There were a total of 17 polymorphic sites among the localities, nine in the ND-5 fragment and eight in the ND-6 fragment; all 17 were transitional changes. Nucleon and nucleotide diversities within each sample ranged from 0.86 ± 0.09 to 0.91 ± 0.05 and from 3.6 x $10^{-3}$ to 4.5 x $10^{-3}$, respectively. Mean pairwise (sequence) differences paralleled nucleotide diversity, ranging from 1.376 in SR1456 to 2.846 in High Falls.

**Population structure:**

Following Bonferroni correction, significant heterogeneity (exact tests) of both allele and genotype distributions at 10 microsatellites (*Nme* 12D8.125 excluded) among the three samples was detected only for microsatellite *Nme* 5G4.175 ($P_{[GENIC]}$ = 0.000; $P_{[GENOTYPIC]}$ = 0.000). Probability values of "genic" and "genotypic" tests for the remaining nine microsatellites ranged from 0.104 – 0.766 and 0.147 – 0.782, respectively. Pairwise comparisons (exact tests of allele and genotype distributions) of samples at *Nme* 5G4.175 indicated significant differences between Confluence and High Falls ($P_{[GENIC]}$ = 0.002, $P_{[GENOTYPIC]}$ = 0.007) and Confluence and SR1456 ($P_{[GENIC]}$ = 0.000, $P_{[GENOTYPIC]}$ = 0.000). We also found significant heterogeneity between Confluence and SR1456 at microsatellite *Nme* 27F8.230 before but not after Bonferroni correction. Pairwise comparisons for the remaining nine microsatellites were non-significant. These results were similar to those reported by Burridge and Gold (2003) for 11 other nuclear-encoded loci where significant heterogeneity (exact tests) in allele distribution at two microsatellites (*Nme* 24B6.211 and *Nme* 4F4.154) was found before but not after Bonferroni correction.

For all subsequent analyses, we combined the 21 nuclear-encoded loci assayed to date where genotype proportions were in HW equilibrium (i.e., excluding only *Nme* 12D8.125). Fisher's method of combining probabilities from independent (exact) tests of all 21 loci revealed highly significant differences (P = 0.000) among samples in global tests of both allele and genotype distributions. Fisher's method when employed in pairwise (exact) tests, revealed significant differences between Confluence and High Falls ($P_{[GENIC]}$ = 0.004, $P_{[GENOTYPIC]}$ = 0.005) and between Confluence and SR1456 ($P_{[GENIC]}$ = 0.000, $P_{[GENOTYPIC]}$ = 0.000); comparisons between High Falls and SR1456 were non-significant ($P_{[GENIC]}$ = 0.540, $P_{[GENOTYPIC]}$ = 0.483). Analysis of molecular variance (Amova) across all loci indicated that between 1.5% (iam) and 2.8% (smm) of the molecular variance was distributed among samples. Corresponding $\Phi_{ST}$ values differed significantly from zero: iam ($\Phi_{ST}$ = 0.015, P = 0.007) and smm ($\Phi_{ST}$ = 0.028, P = 0.038). Locus-by-locus Amova

yielded similar results: IAM (average $\Phi_{ST}$ = 0.014, P = 0.006) and SMM (average $\Phi_{ST}$ = 0.015, P = 0.006).

An exact (global) test of homogeneity in mtDNA haplotype distribution among samples was non-significant (P = 0.835), as were pairwise exact tests (0.642 < P < 0.846). Similarly, AMOVA revealed that 100% of the molecular variation in mtDNA was distributed within samples, with a non-significant (P = 0.648) $\Phi_{ST}$ value of - 0.017.

**Effective population size**:

Results of "*M*" tests are shown in Table 3. *M* is the mean ratio of the number of alleles to the range in allele size, $M_c$ is the critical value of *M* under a specified mutation model (i.e., the "cutoff" above which 95% of simulated values of *M* occur under a hypothesis of mutation-drift equilibrium), and P represents the probability of obtaining the observed value of *M* in an equilibrium population. *M* values for Confluence and High Falls are significantly below $M_c$ indicating that both have experienced recent reductions in effective size and are not in mutation/drift equilibrium; the *M* value for SR1456 is of borderline significance.

Estimates of long-term or inbreeding effective size ($N_{eI}$) and 95% confidence intervals (in parentheses) for the three samples were as follows: Confluence - 2,063 (1,884–2,267); High Falls - 1,314 (1,197–1,460); and SR1456 - 2,972 (2,729–3,248). Based on the 95% confidence intervals, all three estimates of $N_{eI}$ differ significantly from one another.

## DISCUSSION

Results of this study are consistent with those of Burridge and Gold (2003) in that Cape Fear shiners from the localities sampled do not appear genetically impoverished. From 6.2 to 7.9 alleles, on average, were found at the 11 microsatellites, and expected heterozygosity (gene diversity) values ranged from 0.648 to 0.889. By comparison, the average number of alleles at five microsatellites among several samples of spotted sunfish (*Lepomis punctatus*) was 7.4 (McElroy et al. 2003), while the average number of alleles at seven of 11 microsatellites among samples of chinook salmon (*Oncorhynchus tshawytscha)* was ~8.0 (Kinnison et al. 2002). In both of these latter two studies, sample sizes

Table 3. Results of tests to detect reductions in population size: *M* is the mean ratio of allele number to range in allele size, $M_c$ is the critical value of *M* based on a specified mutation model, and Probability is the percentage of time a smaller $M/M_c$ ratio would be expected under equilibrium conditions.

| Sample locality | *M* | $M_C$ | Probability |
|---|---|---|---|
| Confluence | 0.771 | 0.814 | 0.008 |
| High Falls | 0.800 | 0.826 | 0.018 |
| SR1456 | 0.798 | 0.776 | 0.103 |

per locality far exceeded those of our study of Cape Fear shiners. DeWoody and Avise (2000) reported that the average heterozygosity for freshwater fishes (> 7,500 individuals, 75 microsatellites) was approximately 0.54, lower than the corresponding values observed in Cape Fear shiners. MtDNA variation (nucleon diversity range of 0.86–0.91) in Cape Fear shiners also was commensurate with that reported in other fishes (Gold and Richardson 1998) and higher than that observed in two other, endangered North American cyprinids (Garrigan et al. 2002). Given that allele diversity is a sensitive indicator of recent and severe bottlenecks (Spencer et al. 2000), we believe it unlikely that Cape Fear shiners are in danger of immediate extinction because of abnormally low levels of genetic variation (for example as in cheetahs: O'Brien et al. 1983, 1985, but see Caro and Laurenson 1994).

Genetic differences among Cape Fear shiners at the localities sampled may suggest limited gene flow and that separate populations exist within the drainage. Results of Amova and of Fisher's method of combining probabilities obtained in exact tests of each of 21 nuclear-encoded loci indicated significant heterogeneity among the three samples, and moreover, that the Confluence locality differed significantly from the other two. Most of the genetic differences appeared to stem from variation at microsatellite *Nme* 5G4.175, although marginal differences were detected in this study at microsatellite *Nme* 27F8.239 and by Burridge and Gold (2003) at microsatellites *Nme* 24B6.211 and *Nme* 4F4.154. No differences in frequencies of mtDNA haplotypes among samples were observed.

*M* values for both Confluence and High Falls were significantly below estimates of $M_c$, indicating that populations at both localities are not in mutation/drift equilibrium and have experienced recent reductions in effective size. This result is consistent with the reported declines in abundance of Cape Fear shiners (Pottern and Huish 1986, 1987; A. Howard, pers. comm.) and with the empirical findings (Garza and Williamson 2001) that the value of *M* consistently predicted demographic histories of populations thought to have experienced recent reductions in effective size. In permanently reduced populations, the value of M can at least partially "recover," while allelic diversity does not (Garza and Williamson 2001). This suggests that future sampling and genetic analysis of Cape Fear shiners might be useful in assessing whether these populations had been recently reduced in size or historically were small.

We employed a coalescent-based approach to estimate the long-term (inbreeding) effective size ($N_{eI}$) of the three samples of Cape Fear shiners. Estimates for the three samples ranged from ~1,300 to ~3,000. These values are much lower than analogous estimates (based on mtDNA) for two critically endangered cyprinid species

(humpback chub and bonytail chub) from the Colorado River drainage, where long-term (female) effective sizes were estimated to be 97,500 and 89,500, respectively (Garrigan et al. 2002). Assuming equal sex ratios, the estimates for the Colorado River cyprinids are approximately two orders of magnitude greater than estimates for Cape Fear shiners. This may indicate that effective sizes of Cape Fear shiners have been small historically, or at least since the last significant bottleneck event. It also may suggest that Cape Fear shiners never have been particularly abundant, as suggested by Snelson (1971). However, long-term, effective sizes of ~1,000 to ~3,000 are well above the "benchmark" estimate of 500 theoretically needed to maintain the evolutionary potential of a population (Franklin 1980). Given that effective size reflects in part the capacity for adaptive response to environmental and/or ecological change (Lande and Barrowclough 1987), the historical effective sizes in Cape Fear shiners appear to have been sufficient to offset extinction due to genetic factors. Interestingly, the 95% confidence intervals for the estimates of effective size suggested that $N_{eI}$ values differed among the three samples. Coalescent-based estimators of effective size have relatively low variance and are thought to be relatively robust even when sample sizes are not large (Felsenstein 2003). This finding supports the hypothesis that historically there may have been environmental/ecological differences among the three localities that resulted in differences in long-term reproductive success.

## ACKNOWLEDGMENTS

We thank D. Rabon for advice, support, and comments on a draft of the manuscript, S. Karlsson for comments on the manuscript, K. Ross for providing statistics on the comparison of *N. mekistocholas* and *C. carpio* mtDNA sequences, and S. Vellas and F. Dang for assistance with parallel-processing of MIGRATE on the SGI Origin 3800. Work was supported by U.S. Fish and Wildlife Service (Department of the Interior) Award 1448-40181-00-G-119, and by Texas Agricultural Experiment Station Project H-6703. The paper was written while the first author was on developmental leave under an award from the U.S.-Norway Fulbright Foundation. This paper represents Contribution Number 119 of the Center for Biosystematics and Biodiversity at Texas A&M University.

## LITERATURE CITED

Alderman, J.M. 1995. Cape Fear Shiner. Pp. 18–20, *In* North Carolina Wildlife Resources Commission Annual Performance Report, Non-game and Endangered Wildlife Program, Volume IV. Project No. E-3, Segment No. 12. Raleigh, NC.

Avise, J.C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA. 464 pp.

Beerli, P. 2002. MIGRATE: Documentation and program, part of LAMARC. Version 1.5. Available at http://evolution.genetics.washington.edu/lamarc.html.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences 98:4563–4568.

Burridge, C.P., and J.R. Gold. 2003. Conservation genetic studies of the endangered Cape Fear shiner, *Notropis mekistocholas* (Teleostei: Cyprinidae). Conservation Genetics 4:219–225.

Caro, T.M., and M.K. Laurenson. 1994. Ecological and genetic factors in conservation: A cautionary tale. Science 263:485–486.

Cornuet, J.M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014.

DeWoody, J.A., and J.C. Avise. 2000. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. Journal of Fish Biology 56:461–473.

El Mousadik, A., and R.J. Petit. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theoretical and Applied Genetics 92:832–839.

Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial restriction data. Genetics 131:479–491.

Felsenstein, J. 2003. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, MA. 580 pp.

Franklin, I.I. 1980. Evolutionary change in small populations. Pp. 135–150, *In* M.E. Soulé, and B.A. Wilcox (Eds.). Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer Associates, Inc., Sunderland, MA. 395 pp.

Garrigan, D., P.C. March, and T E. Dowling. 2002. Long-term effective population size of three endangered Colorado River fishes. Animal Conservation 5:95–102.

Garza, J.C., and E.G. Williamson. 2001. Detection of reduction in population size using data from microsatellite loci. Molecular Ecology 10:305–318.

Gold J.R., and L.R. Richardson. 1998. Mitochondrial DNA diversification and population structure in fishes from the Gulf of Mexico and western Atlantic. Journal of Heredity 89:404–414.

Heist, E.J., and J.R. Gold. 2000. DNA microsatellite loci and genetic structure of red snapper (*Lutjanus campechanus*) in the Gulf of Mexico. Transactions of the American Fisheries Society 129:469–475.

Hilton-Taylor, C. 2000. 2000 IUCN Red List of Threatened Species. IUCN, Cambridge, UK.

Howard, A., T. Kwak, and G. Cope. 2002. What limits the endangered Cape Fear shiner: Water quality or physical habitat? Proceeding of the 2002 Annual Meeting of the American Society of Ichthyology and Herpetology, Kansas City, MO.

Jarne, P., and P.J.L. Lagoda. 1996. Microsatellites: From molecules to populations and back. Trends in Ecology and Evolution 11:424–429.

Kingman, J.F.C. 1982. The coalescent. Stochastic Processes and their Application 13:235–248.

Kinnison, M.T., P. Bentzen, M.J. Unwin, and T.P. Quinn. 2002. Reconstructing recent divergence: Evaluating nonequilibrium population structure in New Zealand chinook salmon. Molecular Ecology 11:739–754.

Lande, R., and G.F. Barrowclough. 1987. Effective population size, genetic variation, and their use in population management. Pp. 87–123, *In* M.E. Soulé (Ed.). Viable Populations for Conservation. Cambridge University Press, Cambridge, UK. 206 pp.

McElroy, T.C., K.L. Kandl, J. Garcia, and J.L. Trexler. 2003. Extinction-colonization dynamics structure genetic variation of spotted sunfish (*Lepomis punctatus*) in the Florida Everglades. Molecular Ecology 12:355–368.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, NY. 512 pp.

O'Brien, S.J., D.E. Wildt, D. Goldman, C.R. Merril, and M. Bush. 1983. The cheetah is depauperate in genetic variation. Science 221:459–462.

O'Brien, S.J., M.E. Roelke, L. Marker, A. Newman, C.A. Winkler, D. Meltzer, L. Colly, J.F. Evermann, M. Bush, and D.E. Wildt. 1985. Genetic basis for species vulnerability in the cheetah. Science 227:1428–1434.

Pottern, G.B., and M.T. Huish. 1986. Supplement to the status survey of the Cape Fear shiner *Notropis mekistocholas*. U.S. Fish and Wildlife Service, Endangered Species Office, Asheville, NC.

Pottern, G.B., and M.T. Huish. 1987. Second supplement to the status survey of the Cape Fear shiner *Notropis mekistocholas*. U.S. Fish and Wildlife Service, Endangered Species Office, Asheville, NC.

Raymond, M, and F. Rousset. 1995. GENEPOP, Version 1.2: Population genetics software for exact test ecumenism. Journal of Heredity 86:248–249.

Schneider, S., D. Roessli, and L. Excoffier. 2000. ARLEQUIN: A software for population genetics data analysis, ver 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland. 111 pp.

Snelson, F.F. 1971. *Notropis mekistocholas*, a new herbivorous cyprinid fish endemic to the Cape Fear River basin, North Carolina. Copeia 1971:449–462.

Spencer, C.C., J.E. Neigel, and P.L. Ledberg. 2000. Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Molecular Ecology 9:1517–1528.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Appendix 1. Repeat sequence, annealing temperature, number of alleles, and PCR primers (forward above, reverse below) for 11 nuclear-encoded microsatellites of *Notropis mekistocholas*. GenBank Accession Numbers for cloned sequences are AY460190 – AY460199.

| Micro-satellite | Repeat sequence | Annealing temperature | # alleles detected | PCR primer sequences (5'⊂3') |
|---|---|---|---|---|
| *Nme* 5B10.211 | $(TG)_{10}$ | 42 | 12 | GAAAGTCCCCATAAAACA |
| | | | | GACCCGTACAGTGCTCT |
| *Nme* 27F8.230 | $(CA)_9$ | 48 | 8 | ACTAAACATACTGCAAAACATACAA |
| | | | | GAGAGGGGGAAAGAGTTG |
| *Nme* 12D8.125 | $(GT)_{10}$ | 51 | 9 | CGGACGCGCAAACATTA |
| | | | | GAACCTCGTCCCTCAAAAA |
| *Nme* 15F2.174 | $(AC)_{10}$ | 58 | 7 | AATGGTATTATCCCCTCCTGACACTT |
| | | | | ATCGGAATAAAGAATTGAATACTGGTTG |
| *Nme* 30F12.208 | $(AC)_{13}$ | 49 | 10 | TACATCATGGCCCTAACACA |
| | | | | GGGCTAAAATTTGGACGAA |
| *Nme* 33B6.125 | $(GT)_8$ | 45 | 6 | TTCTGGGATAAATCACAACTT |
| | | | | TGGGCCATAATCACTCAT |
| *Nme* 2B10.166* | $(TG)_9$ | 59 | 6 | CAGAAGTGCCAGCGAGAATGT |
| | | | | TCAAAGATGCCCGCCCCTAATA |
| *Nme* 2B10.232* | $(CA)_9$ | 54 | 14 | ATGGTCATGGCCTAAAGCAATAC |
| | | | | AATGCAAAAGCAACTGTAATCAACTA |
| *Nme* 5G4.175 | $(CT)_{13}$ | 47 | 11 | AAGCATTACCCCGTTTTG |
| | | | | GACAGGTGGACATTGAATCA |
| *Nme* 6A7.93 | $(CA)_{13}$ | 52 | 12 | CACCAAACTGTCATTCAAATAAG |
| | | | | GACCCTGGGCGTTCTCTG |
| *Nme* 2D5.117 | $(TG)_{10}$ | 56 | 11 | CAGATTGAGCCGAACGATTTG |
| | | | | TTTTTGCCCAGCCCTAGAATACAC |

* Microsatellites *Nme* 2B10.166 and *Nme* 2B10.232 were designed from the same cloned fragment.