Conservation genetics and demographic history of the endangered Cape Fear shiner (Notropis mekistocholas)

E. SAILLANT, J. C. PATTON, K. E. ROSS and J. R. GOLD
Center for Biosystematics and Biodiversity, Texas A&M University, College Station, Texas, 77843–2258, USA

Abstract

We examined allelic variation at 22 nuclear-encoded markers (21 microsatellites and one anonymous locus) and mitochondrial (mt)DNA in two geographical samples of the endangered cyprinid fish Notropis mekistocholas (Cape Fear shiner). Genetic diversity was relatively high in comparison to other endangered vertebrates, and there was no evidence of small population effects despite the low abundance reported for the species. Significant heterogeneity (following Bonferroni correction) in allele distribution at three microsatellites and in haplotype distribution in mtDNA was detected between the two localities. This heterogeneity may be due to reduced gene flow caused by a dam built in the early 1900s. Bayesian coalescent analysis of microsatellite variation indicated that effective population size of Cape Fear shiners has declined in recent times (11–25 435 years ago, with highest posterior probabilities between 126 and 2007 years ago) by one–two orders of magnitude, consistent with the observed decline in abundance of the species. A decline in effective size was not indicated by analysis of mtDNA, where sequence polymorphism appeared to carry the signature of an older expansion phase that dated to the Pleistocene (~12 700 > 1 million years ago). Cape Fear shiners thus appear to have undergone an expansion phase following a glacial cycle but to have declined significantly in more recent times. These results suggest that rapidly evolving markers such as microsatellites may constitute a suitable tool when inferring recent demographic dynamics of populations.

Keywords: Cape Fear shiners, conservation genetics, demographic history, effective population size, microsatellites, mtDNA

Received 19 March 2004; revision received 7 June 2004; accepted 2 July 2004

Introduction

Efforts to manage and conserve threatened and endangered populations of freshwater fishes have been supported in many ways by deployment of molecular markers to resolve taxonomic problems and describe patterns of genetic variations both within and among lineages (Vrijenhoek 1998) and to infer evolutionary and demographic history of species under decline (Moran 2002). The latter has been improved greatly by the recent advent of genealogical modelling of the coalescent process, permitting comparisons of statistics calculated from a data sample with the distribution of the statistics in simulated samples (Beaumont 1999). Coalescent-based analyses thus allow, in theory, robust inference even when limited sample sizes are available, as it is often the case for endangered species.

In this study, we utilized a suite of molecular markers to describe patterns of genetic variation and infer demographic history of the Cape Fear shiner, Notropis mekistocholas (Snelson 1971), a small, herbivorous cyprinid fish (minnow) endemic to the Cape Fear river basin in the eastern Piedmont region of North Carolina (USA) and one of the 187 species of freshwater and diadromous fishes in the southern USA recognized as extinct, endangered or threatened (Warren et al. 2000). Cape Fear shiners are reported to occur in only five localities in the Cape Fear drainage (North Carolina Wildlife Resources Commission 1995), but in recent years have not been found in two of these localities, both in the Haw River (D. Rabon, US Fish and Wildlife Service, personal communication). In addition, very recent surveys at a third locality, the Deep River above the dam at High Falls in Moore County, have...
yielded only a single Cape Fear shiner (A. Howard, North Carolina State University, personal communication). While it may be that Cape Fear shiners have always existed in low numbers (Snelson 1971), Pottern & Huish (1986, 1987) reported declines in abundance and range for the species that were not evident for other sympatric taxa. Cape Fear shiners are currently designated as endangered (Federal Register 1987; Hilton-Taylor 2000; http://www.unep-wcmc.org/index.html? http://www.unep-wcmc.org/species/index.htm–main); potential causes for their decline include changes in stream morphology and flow (US Fish and Wildlife Service 1988) and reduced habitat quality and availability (Howard 2003).

To date, allelic variation at 23 genetic markers (21 microsatellites, one anonymous nuclear-encoded locus and mtDNA) has been examined among samples collected in 1998 from the three localities where Cape Fear shiners have been captured in the last decade (Burr ridge & Gold 2003; Gold et al. 2004). Although relatively high levels of genetic diversity and no evidence of small population effects or inbreeding were found, a recent reduction in effective population size was suggested by the ratio of the number of (microsatellite) alleles to the range in allele size (the $M$-test of Garza & Williamson 2001). However, sample sizes employed in these studies were small (10–15 individuals per locality), due in large part to the rarity and difficulty in capturing Cape Fear shiners. Here, we report results from (noninvasive) sampling of additional specimens at the two localities where Cape Fear shiners still occur in relative abundance. We assayed the same set of 23 genetic markers used by Gold et al. (2004). Our objectives were to (i) further examine genetic diversity and effective population size in extant Cape Fear shiners in order to ascertain whether small population effects were evident; (ii) assess more rigorously whether genetic heterogeneity existed between the sample localities; and (iii) employ recently developed, coalescent approaches (Kuhner et al. 1998; Beaumont 1999; Storz & Beaumont 2002) and mismatch distributions (Rogers & Harpending 1992) to infer the demographic dynamics of Cape Fear shiners at each locality.

Materials and methods

Study sites and sample collection

Adult Cape Fear shiners were sampled during the spring and summer of 2002 from two of the three localities sampled in 1998. Full details of the sampling localities may be found in Burr ridge & Gold (2003); a map of the area, including sampling localities, is shown in Fig. 1. The two localities sampled in 2002 were Confluence ($n = 29$) and High Falls ($n = 26$); Confluence is located 13.5 miles downstream from High Falls and the two localities are separated by the Carbonton Dam built in 1922 (D. Rabon, US Fish and Wildlife Service, personal communication). Small clips from the lower lobe of the caudal fin were removed, stored in 95% ethanol and returned to College Station, where they were stored at $-80\,^\circ C$. All the fin-clipped fish swam away with no noticeable effect. We also included in this analysis the 15 samples from both Confluence and High Falls obtained in 1998; thus the overall sample sizes for Confluence and High Falls were 44 and 41, respectively.

Genetic assays

All specimens were genotyped at 22 nuclear-encoded genetic markers. These markers included 21 microsatellites and one anonymous locus (Nme18A6.0) that coamplifies with microsatellite Nme18A6.158. Details of polymerase chain reaction (PCR) amplification, electrophoresis, and scoring of PCR products may be found in Burr ridge & Gold (2003) and Gold et al. (2004). Variation in mtDNA was assessed by directly sequencing fragments within the ND-5 [258 base pairs (bp)] and ND-6 (367 bp) subunits of mitochondrially encoded NADH dehydrogenase. PCR primer pairs, amplification conditions and sequence alignment procedures may be found in Gold et al. (2004). The partition test of Farris et al. (1995a,b), as implemented in PAUP 4.0b10 (Swofford 1998), was used to assess whether evolutionary rates of the two fragments of mtDNA were
homogeneous. The proportion of randomly generated partitions (100 replicates) that showed a greater incongruence index (Mickevich & Farris 1981) than the sequenced fragments was 0.52; consequently, the hypothesis of congruence was accepted and the sequences from both mtDNA fragments were combined into unique haplotypes for all subsequent analyses.

**Data analysis**

Allele (nuclear) and haplotype (mtDNA) frequencies, number of alleles and haplotypes, allele richness (a measure of the number of alleles independent of sample size) and unbiased gene (microsatellites) diversity were computed using FSTAT version 2.9.3 (Goudet 1995). MtDNA number of polymorphic sites and nucleon and nucleotide diversity were computed in ARLEQUIN (Schneider et al. 2000). Gene and nucleon diversity were estimated after Nei (1987). Homogeneity of allele richness and of gene diversity between pairs of samples was tested using Wilcoxon signed-rank tests. Conformity of genotype proportions to Hardy–Weinberg equilibrium expectations within samples was measured for each nuclear-encoded locus as Weir & Cockerham (1984) test. Conformity of genotype proportions from Hardy–Weinberg expectations was assessed via an exact test, as implemented in GENEPOL. The means of the prior distributions of the variance of \( \theta \) among microsatellites for these three parameters were set to 0.5 so that expected ratios for pairs of microsatellites would be of approximately fivefold under the prior (Storz & Beaumont 2002). Mean and SD for the prior distribution of \( \mu \) were set to \( -3.3 \) and 0.5 so that values in the range \( 5 \times 10^{-3} \)–\( 5 \times 10^{-5} \) would have reasonable support (Storz & Beaumont 2002); this prior thus utilized values that are widely assumed in demographic analyses (Heath et al. 2002; Storz & Beaumont 2002). The SD of the prior distribution of the variance among microsatellites was set to 2 so that an expected ratio of mutation rates for pairs of microsatellites of about 700-fold would be frequent under the prior (Storz & Beaumont 2002). The generation time of Cape Fear shiners is not known; however, life-history data on other shiners (Harrell & Cloutman 1978; Matthews & Heins 1984; Cloutman & Harrell 1987) indicate that a generation time between 1 and 3 years would be appropriate. Consequently, time since expansion/decline was estimated considering generation time between 1 and 3 years. Estimates of current effective population size \( N_e \) and confidence limits were calculated by using the relationship \( N_e = N_0/2 \).

**Demographic history**

The demographic history of Cape Fear shiners at each locality was explored using microsatellite data and a Bayesian coalescent-based approach (Beaumont 1999; Storz & Beaumont 2002). A population changing in size exponentially was considered, and demographic parameters were inferred based on microsatellite genotypes and assuming a stepwise mutation model. The parameters estimated were \( N_0 \), \( N_1 \), \( \mu \) and \( t_e \). \( N_0 \) and \( N_1 \) are the effective number of chromosomes at sampling and at the beginning of an expansion/decline phase, respectively; \( \mu \) is the average mutation rate over loci per generation; and \( t_e \) is the time since the beginning of an expansion/decline phase. The ratio \( r = N_0/N_1 \) is expected to be \( < 1 \) in a declining population, equal to 1 in a stable population and \( > 1 \) in an expanding population (Beaumont 1999). Monte Carlo Markov chain (MCMC) simulations were used to estimate the posterior distribution of the genealogical history (mutational and coalescent events) and demographic parameters (growth rate and time since expansion) for each locality. Three chains (10⁸ steps per chain, with the first 5 × 10⁸ steps discarded to remove possible bias due to the starting values of the parameters) were run using the program MSVAR1.3.exe, kindly provided by M. Beaumont of the University of Reading, UK. The three chains were based on different starting parameters in order to confirm convergence of the MCMC. The means of the prior distributions of means \( N_0 \), \( N_1 \) and \( t_e \) were set (on a log₁₀ scale) to 4 and their standard deviations (SD) were set to 3 in order to provide support for a wide a priori range for the parameters. The SDs of the prior distributions of the variance among microsatellites for these three parameters were set to 0.5 so that expected ratios for pairs of microsatellites would be of approximately fivefold under the prior (Storz & Beaumont 2002). Mean and SD for the prior distribution of \( \mu \) were set to \( -3.3 \) and 0.5 so that values in the range \( 5 \times 10^{-3} \)–\( 5 \times 10^{-5} \) would have reasonable support (Storz & Beaumont 2002); this prior thus utilized values that are widely assumed in demographic analyses (Heath et al. 2002; Storz & Beaumont 2002). The SD of the prior distribution of the variance of \( \mu \) among microsatellites was set to 2 so that an expected ratio of mutation rates for pairs of microsatellites of about 700-fold would be frequent under the prior (Storz & Beaumont 2002). The generation time of Cape Fear shiners is not known; however, life-history data on other shiners (Harrell & Cloutman 1978; Matthews & Heins 1984; Cloutman & Harrell 1987) indicate that a generation time between 1 and 3 years would be appropriate. Consequently, time since expansion/decline was estimated considering generation time between 1 and 3 years. Estimates of current effective population size \( N_e \) and confidence limits were calculated by using the relationship \( N_e = N_0/2 \).
The mtDNA haplotype sequence data also were used to infer demographic history at each locality. Neutrality of the mtDNA sequences was tested by means of Fu’s (1997) Fs statistic, as implemented in ARLEQUIN, and Fu & Li’s (1993) D* and F* statistics, as implemented in DNASP (Rozas et al. 2003). Significance of Fs was assessed by 10 000 randomizations; significance of D* and F* was assessed using 10 000 coalescent simulations based on the observed number of segregating sites in each sample. Demographic history of mtDNA was first examined using the maximum-likelihood coalescent approach of Kuhner et al. (1998), as implemented in the program FLUCTUATE version 1.4, available at http://evolution.genetics.washington.edu/lamarc/fluactuate.html. FLUCTUATE utilizes a model where a single population has been expanding (or declining) exponentially and provides estimates of the parameters Θ and g. Θ is defined as 4Neµ (where Ne is the ‘current’ effective population size and µ is the neutral mutation rate per site), and g is the exponential growth rate of the population (in 1/µ generations). Watterson (1975) segregating sites estimate was used as the initial estimate of Θ for each run. FLUCTUATE requires input of a transition/transversion ratio to be applied during estimation. To determine this ratio, the model of DNA substitution that best fitted the mtDNA sequence data was determined based on the likelihood-ratio test statistic, as implemented in MODELTEST version 3.06 (Posada & Crandall 1998). The best model [H–Kishino–Yano (HKY) model of Hasegawa et al. 1985] involved variable base frequencies and a transition/transversion ratio of 13 for the mtDNA sequences assayed. This ratio was used to parameterize the mutation model implemented by FLUCTUATE. One hundred short chains (1000 steps per chain with a sampling increment of 100) followed by 10 long chains (500 000 steps per chain with a sampling increment of 500) were employed to ensure convergence of the MCMC. Estimates of time (in generations) since a ‘population’ differed by one and two orders of magnitude from the ‘current’ effective size were determined by using the formula \( t_e = (\ln(N_e/N_i))/(g\mu) \), where \( N_i \) and \( N_e \) are the effective population sizes at present (\( N_i \)) and \( t \) generations ago (\( N_e \)), respectively. Given the wide range of mutation rates that have been estimated for coding regions of mtDNA in fish (e.g., 1.2% for NADH2 in several gmate species pairs of fishes according to Bermingham et al. 1997 to c. 5% for ND5/6 in salmonids according to Doiron et al. 2002), values of µ employed in the estimation were 1.0 and 5.0% MY\(^{-1}\). Demographic history based on mtDNA sequence variation was then examined by mismatch-distributions analysis (Rogers & Harpending 1992). The sum-of-squared-difference statistic (SSD) was used to compare the observed (mismatch) distribution of pairwise differences between individuals to that expected under demographic expansion; significance of SSD was assessed by parametric bootstrap resampling (Schneider & Excoffier 1999), as implemented in ARLEQUIN (10 000 resamplings).

**Results**

Detailed summary statistics, including allele and genotype distributions for the 22 nuclear-encoded loci and haplotype distributions for mtDNA, for each locality and each year sampled may be obtained from the corresponding author. Temporal variation, i.e. between sample years, in allele frequency over all 22 (21 microsatellites) nuclear-encoded loci was nonsignificant in both Confluence (\( P = 0.176 \)) and High Falls (\( P = 0.790 \)). Probability values for individual loci ranged from 0.011 (Nm30D2-120 in Confluence) to unity (Nm30D2-120 and Nm24B6-191 in High Falls); none of the probability values were significant following Bonferroni correction. MtDNA haplotype frequencies also did not differ significantly between sample years (\( P = 0.134 \), Confluence; \( P = 0.835 \), High Falls).

Summary data of variation for each nuclear-encoded locus at each sample locality, pooled across years, are presented in Appendix I. All nuclear-encoded loci were polymorphic. Number of alleles sampled per locus per locality ranged from two (Nm24B6.191, both localities; Nm30D2.120, High Falls) to 14 (Nm30F12.208, both localities); allelic richness per locus averaged 5.5 and ranged from 1.76 (Nm24B6.191, Confluence) to 8.4 (Nm2B10.232, Confluence). Gene diversity per locus averaged 0.703 and ranged from 0.128 (Nm24B6.191, Confluence) to 0.889 (Nm2D5.117, High Falls). No significant difference in allelic richness (\( P = 0.81 \)) or gene diversity (\( P = 0.62 \)) was detected in pairwise comparisons between localities. Tests of conformity to Hardy–Weinberg equilibrium expectations were nonsignificant following Bonferroni correction except for Nm12D8.125, where a highly significant deficit in heterozygotes was observed for both localities (\( F_{IS} \) values were 0.372 and 0.672, respectively, for Confluence and High Falls). The heterozygote deficiency may reflect occurrence of null alleles; consequently, Nm12D8.125 was omitted from all subsequent analyses. A total of 22 of 462 (pairwise) tests of genotypic disequilibrium were significant (\( P < 0.05 \)) prior to Bonferroni correction; of these, two (Nm2B10.166/ Nm2B10.232, Confluence; and Nm27F8.230/Nm2B10.232, High Falls) remained significant after correction. A total of 19 distinct mtDNA haplotypes were detected (Fig. 2). Nucleotide diversity values were 0.80 ± 0.06 (Confluence) and 0.85 ± 0.03 (High Falls), and nucleotide diversity values were 3.6 × 10\(^{-3} \) (Confluence) and 4.3 × 10\(^{-3} \) (High Falls).

**Geographic structure**

Significant heterogeneity (\( P < 0.05 \)) of allele frequencies between the two localities was detected at 14 of the...
Fig. 2 Network of haplotypes of Cape Fear shiners (Notropis mekistocholas) sampled within the Cape Fear river drainage. Haplotypes (H) found in Confluence are in the white area (overall sample size n = 44); those in the black area are for High Falls (overall sample size n = 41). • missing haplotype. GenBank Accession nos for composite haplotypes are as follows (ND5–ND6): H1 (AY396545–AY396559), H2 (AY396546–AY396555), H3 (AY396547–AY396556), H4 (AY396548–AY396556), H5 (AY396545–AY396557), H6 (AY396549–AY396558), H7 (AY396550–AY396559), H8 (AY396551–AY396560), H9 (AY396552–AY396556), H10 (AY396545–AY396561), H11 (AY396553–AY396556), H12 (AY396554–AY396560), H13 (AY422787–AY422791), H14 (AY396547–AY422792), H15 (AY422787–AY396555), H16 (AY396545–AY422792), H17 (AY422787–AY396559), H18 (AY422790–AY396562), H19 (AY396545–AY422793).

nuclear-encoded loci prior to Bonferroni correction; three of these (Nme2B10.232, Nme5G4.175, and Nme6A7.93) remained significant after correction. The overall degree of divergence was relatively low (Fst of 0.014, P < 0.0001). Significant heterogeneity (FST = 0.029, P = 0.035) between localities in mtDNA haplotype frequencies also was detected. The mtDNA haplotype topology (Fig. 2) revealed no evident phylogeographical separation (reciprocal monophyly) of haplotypes within either locality.

Demographic history

The anonymous, nuclear-encoded locus (Nme18A6.0) that complies with Nme18A6.158 was removed from this analysis, as variation at this locus could not be assumed to follow the stepwise mutation model assumed in msvar computations. In addition, alleles that differed by only a single base pair were detected at five of the (20) remaining loci. Such alleles were very infrequent at three of the loci (one at Nme18A6.158 and Nme5B10.211, and three at Nme24B6.211). Genotypes of the corresponding individuals (total of four) at these three loci were removed from the data set and only genotypes of the remaining individuals at each locus were used in the analysis. A total of 19 and 84 such alleles were found at microsatellites Nme7B9.190 and Nme27F8.230, respectively. These two microsatellites were omitted from the analysis; the final data set thus included 18 microsatellites.

Bayesian coalescent analysis revealed a negative posterior distribution of log10(r) values, consistent with a historical decline in effective size among Cape Fear shiners at both localities (Table 1). The corresponding (estimated) reduction in effective size ranged from one to three orders of magnitude, with modes between 0.97 (Confluence) and 2.08 (High Falls) (Table 1). Estimates of Ne (the current effective population size) ranged from 7 to 4463, with modes at 513 (Confluence) and 85 (High Falls). Corresponding estimated average mutation rates for the 18 microsatellites analysed were between 4.5 × 10⁻⁵ and 1.7 × 10⁻³ (mode = 2.5 × 10⁻⁴). Given the assumed generation-time range of 1–3 years, the estimated time since decline for Confluence ranged from 62 to 25 435 years ago (mode between 669 and 2007 years ago, depending on the estimate of generation time); for High Falls the time since decline ranged from 11 to 6323 years ago (mode between 126 and 379 years ago).

Results of neutrality tests for mtDNA sequences are shown in Table 2. Fu’s Fs statistic indicated a significant departure from neutrality for the sample from Confluence (P = 0.007) but not for the sample from High Falls (P = 0.200). Fu & Li’s D* and F* statistics were nonsignificant (P > 0.05) for both samples. According to Fu (1997), these results indicate that population expansion or ‘hitchhiking’, rather than background selection, accounts for the significant departure from neutrality detected in the Confluence sample by Fs. The hypothesis of population expansion under the assumption of selective neutrality was examined further by the coalescent approach of Kuhner et al. (1998) and by mismatch-distribution analysis (Rogers

<table>
<thead>
<tr>
<th></th>
<th>Mode</th>
<th>0.05 quartile</th>
<th>0.95 quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confluence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>μ</td>
<td>2.5 × 10⁻⁴</td>
<td>4.5 × 10⁻⁵</td>
<td>1.7 × 10⁻³</td>
</tr>
<tr>
<td>Nₑ</td>
<td>513</td>
<td>54</td>
<td>4463</td>
</tr>
<tr>
<td>Log₁₀(r)</td>
<td>−0.97</td>
<td>−0.72</td>
<td>−1.89</td>
</tr>
</tbody>
</table>

| **High Falls**   |         |              |              |
| μ                | 2.5 × 10⁻⁴ | 4.2 × 10⁻⁵   | 1.6 × 10⁻³   |
| Nₑ              | 85      | 7            | 1235         |
| Log₁₀(r)        | −2.08   | −1.20        | −2.87        |
| tₑ (year)†      | 126–379 | 2108–6323    | 11–33        |

†Range covering generation times between 1 and 3 years.
Table 2  Fu's (1997) $F_s$, and Fu & Li's (1993) $D^*$ and $F^*$ measures of selective neutrality; probabilities of significance were estimated from randomizations ($F_s$) or coalescent simulations ($D^*$ and $F^*$)

<table>
<thead>
<tr>
<th></th>
<th>$F_s$</th>
<th>$P(F_s &lt; F_{s\text{obs}})$</th>
<th>$F^*$</th>
<th>$P(F^* &lt; F_{*\text{obs}})$</th>
<th>$D^*$</th>
<th>$P(D^* &lt; D_{\text{obs}})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>-5.60</td>
<td>0.007</td>
<td>-1.59</td>
<td>0.079</td>
<td>-1.17</td>
<td>0.099</td>
</tr>
<tr>
<td>High Falls</td>
<td>-1.99</td>
<td>0.200</td>
<td>-0.73</td>
<td>0.244</td>
<td>-0.47</td>
<td>0.379</td>
</tr>
</tbody>
</table>

Fig. 3  Plot of the ratio of historical to present-day effective population size, based on coalescent simulations carried out with mitochondrial DNA sequence data for Cape Fear shiners from two localities within the Cape Fear River drainage (Confluence, $n = 44$, triangles; High Falls, $n = 40$, circles). Per site mutation rates assumed were 1.0% (open triangles and circles) and 5% (filled in triangles and circles) per million years.

Fig. 4  Mismatch distributions based on mitochondrial DNA sequences of Cape Fear shiners from two localities within the Cape Fear River drainage. Bars represent observed frequency of number of pairwise differences between sequences; lines represent the expected distribution assuming demographic expansion. Estimated values of $\tau$ (crest of the mismatch distribution) and its upper and lower bounds, generated by the bootstrap percentile method, are given for both samples.

Discussion

The first objective of this study was to further assess genetic diversity in extant Cape Fear shiners in order to ascertain whether small population effects were evident. Small populations often exhibit reduced levels of genetic diversity and increased probability of identity by descent or inbreeding (O'Brien et al. 1996). Reduced genetic diversity may impact negatively a population's capability to respond to future environmental change or challenge, and increased inbreeding can expose deleterious recessives that can lead to reduction in overall fitness (Frankham 1995; Lynch et al. 1995; Frankham et al. 2002). For the two populations

differ from that expected under an expansion model; the sum-of-squared-difference (SSD) statistics were $0.020 (P_{\text{SSD}} = 0.236)$ for Confluence and $0.018 (P_{\text{SSD}} = 0.157)$ for High Falls. The time at which such demographic expansion would have occurred was estimated using the relationship $\tau = 2ut$, where $\tau$ is the crest of a unimodal mismatch distribution, $u$ is the mutation rate per generation of the DNA region under study and $t$ is the time in generations since demographic expansion (Rogers & Harpending 1992). Values of $\tau$ (Fig. 4) were obtained from ARLEQUIN; the parameter $u$ was estimated as $mT \mu$, where $\mu$ is the mutation rate per nucleotide and $mT$ is the number (625) of nucleotides assayed; $\mu$ values were set to 1.0 and 5.0% MY$^{-1}$ as above. Estimates of the time in years since demographic expansion ranged from 92 320 (Confluence, $\mu = 5.0$% MY$^{-1}$) to 522 240 (High Falls, $\mu = 1$% MY$^{-1}$). Considering the upper and lower bounds of $\tau$ (Fig. 4), the time since demographic expansion could vary between 12 704 and 1126 400 years ago.
examined here, the number of alleles over all nuclear-encoded loci averaged 8.18, while gene diversity ($H_e$) per locus averaged 0.701. These averages are nearly the same as or higher than those (7.5 alleles/locus, gene diversity of 0.46) reported by DeWoody & Avisé (2000) for freshwater fishes, and considerably higher than averages for other critically endangered vertebrates, including the North Atlantic right whale (Waldick et al. 2002), Galápagos penguin (Akst et al. 2002), Gilbert’s potoroo (Sinclair et al. 2002) and Delmarva fox squirrel (Lance et al. 2003). In these endangered species, the number of alleles per microsatellite locus ranged from 2.0 to 3.2, and gene diversities ranged from 0.038 to 0.457. Levels of mtDNA diversity in Cape Fear shiners also were relatively high. Nineteen mtDNA haplotypes were uncovered among the 85 individuals surveyed, and nucleon diversity (the probability that any two individuals drawn at random will differ in mtDNA haplotype) in both populations was greater than 0.80. Collectively, these results indicate that levels of genetic diversity in Cape Fear shiners are relatively high, despite current rarity of the species and its reported decline(s) in diversity in Cape Fear shiners are relatively high, despite the fact that Cape Fear shiner populations appear to have undergone significant effective-size reduction in the recent past (see below), and as pointed out by Hauser et al. (2002) changes in genetic (allelic) diversity may be in part a function of the history of population-size reduction and may not always be detectable, even though the population is in, or has experienced, a recent decline. Overall, however, results presented here suggest that high levels of genetic diversity are still present in current Cape Fear shiner populations.

The second objective of the study was to assess more rigorously whether genetic heterogeneity existed among Cape Fear shiners within the drainage. Burridge & Gold (2003) and Gold et al. (2004) reported significant allele-frequency differences at one or more nuclear-encoded loci between samples taken in 1998 at Confluence and High Falls. For most loci, however, the allele-frequency differences were not significant following Bonferroni adjustment. With the increased sample sizes obtained for this study, significant heterogeneity between the two localities was detected both before (14 microsatellites) and after (three microsatellites) Bonferroni adjustment for the nuclear-encoded loci and for mtDNA as well.

Given that Cape Fear shiners at the two localities differ genetically, it then becomes of interest to ask whether the two populations were separated historically or whether the dam constructed between the two localities in the early 1900s could have impeded gene flow sufficiently to allow genetic differences to accumulate. The former would suggest that the two populations were on independent evolutionary trajectories, whereas the latter might reflect the impact of human activities. Historical separation and the effective absence of gene flow between the two localities would be expected, in time, to lead to reciprocal monophyly of mtDNA haplotypes (Avisé et al. 1987). The network of haplotypes of all Cape Fear shiner mtDNA haplotypes, however, revealed no evident phylogeographical separation (reciprocal monophyly) of haplotypes within either locality. We then asked if genetic drift alone could account for genetic differences observed between the two localities given the time period involved (1922 to present). We used the equation (from Hedrick 1999)

$$F = 1 - \left(1 - \frac{1}{2N} \right)^t$$

where $F$ is $F_{ST}$ among idealized populations under an infinite-island model (and assuming no migration), $N_e$ is effective size and $t$ is time in generations. For $F_{ST}$, we used 0.014 (the weighted $F_{ST}$ over all loci examined) and for $N_e$ we used the harmonic means of the lower ($N_e = 12$) and upper ($N_e = 2417$) bounds of the posterior distributions of $N_e$ obtained for Confluence and High Falls. Assuming a generation time of 1–3 years, and a starting $F_{ST}$ of 0 (no reproductive isolation before dam construction), it would require 1–3 years (one generation) to generate an $F_{ST}$ of 0.014 at $N_e = 12$, and 70–210 years (70 generations) to generate an $F_{ST}$ of 0.014 at $N_e$ of 2418. These findings together suggest that the genetic differences observed between Cape Fear shiners at the Confluence and High Falls localities may be relatively recent and could have arisen from reduced gene flow caused by the dam built in 1922.

The Bayesian coalescent analysis provided evidence for a decline in effective size of one to two orders of magnitude in the two populations, suggesting that Cape Fear shiners may not have always existed in low numbers as suggested by Snelson (1971). The estimated time of the decline was within the last ~10–25 435 years with highest posterior probability of occurrence 126–2007 years ago. The lower bound of this time interval is compatible with the timing of recent anthropogenic degradation of habitat within the Cape Fear River drainage. During the past century, several dams and hydroelectric plants have been constructed in various parts of the drainage (US Fish and Wildlife Service 1988) and undoubtedly modified flow regimes and habitat quality. Increased sedimentation from soil erosion and agricultural runoff has been particularly noteworthy in the drainage (NCDWQ 2000) and would be expected to impact negatively Cape Fear shiners because of their feeding habits and spawning requirements (Howard 2003). Finally, overall water quality in the drainage has deteriorated...
significantly due to relatively high levels of contaminants such as zinc, lead, DDT and chlordanes (Howard 2003).

There are caveats to the above inferences. The hierarchical model implemented in msvar1.3 assumes a stepwise model of mutation and an enclosed population. Mutations potentially can involve multistep changes in allele size that could create gaps in the distribution of allele sizes and produce a pattern of variation similar to that generated by a population bottleneck (Storz & Beaumont 2002). Similarly, immigration of divergent alleles from a different population could create such gaps if the immigrant alleles are outside the range of allele sizes of the recipient population (Garza & Williamson 2001). However, multistep changes appear to be very infrequent relative to single-step changes, and thus are expected to potentially affect only a small proportion of microsatellites (Storz & Beaumont 2002). Because the hierarchical model tends to give less weight in multilocus data sets to microsatellites displaying atypical allele-size distributions (Storz & Beaumont 2002), deployment of a large number of microsatellites, as in this study, is expected to recover adequately demographic signals inherent in the data. Finally, with regard to immigration, the genetic heterogeneity in both microsatellites and mtDNA detected between Confluence and High Falls indicates that migration between the two populations is very limited or absent. Even if migration is limited, allele sizes at the loci studied are sufficiently similar, such that allele-size gaps would not be expected to occur.

Both the coalescent-based analysis of Kuhner et al. (1998) and the mismatch distributions of pairwise differences among mtDNA haplotypes were consistent with population expansion at both localities. The estimated time since expansion based on the coalescent-based analysis was between ~68 500 and 510 000 years ago; similarly, the estimated time since expansion from mismatch-distribution analysis was between ~12 700 and 1.13 million years ago, with highest probability between 92 320 and 522 240 years ago. These times since population expansion, based on mtDNA, are two to three orders of magnitude larger than the estimated time since the population decline detected by variation in the microsatellites. This highlights the different time scales exposed by the two types of genetic markers. The estimate based on mtDNA is consistent with population expansion following a glacial cycle during the Pleistocene, whereas the population decline detected by the microsatellites dates back maximally to the last 25 000 years, and minimally within the last few decades. Similar results were obtained by Lavery et al. (1996), Lessios et al. (2001), and Koskinen et al. (2002), who found that genetic patterns based on mtDNA data were consistent with historical population expansion in species that today are in severe population decline. The results are also consistent with the outcome of a simulation conducted by Rogers (1996); population declines could follow expansion but not affect a genetic signature of expansion in mtDNA polymorphism.

Estimates of \( N_e \) (the current effective population size) ranged from ~7 to ~4463, with highest probability of \( N_e \) of 513 and 85 for Confluence and High Falls, respectively. The minimum effective size needed to ensure long-term integrity is still a matter of debate. A few hundred are, in theory, needed to maintain the equilibrium between loss of adaptive genetic variation due to genetic drift and its replacement by mutation (Franklin 1980; Schultz & Lynch 1997). However, for several reasons Lynch & Lande (1998) argued that the target effective size for a conservation programme for an endangered species should be in the range of 1000–5000. Thus, our estimates of current effective population size in Cape Fear shiner would be close to or lower than these minimum values. Our studies to date indicate that while genetic diversity appears relatively high in the populations of Cape Fear shiners at the two localities, both populations appear to have undergone significant, recent decline in effective size. In addition, effective sizes in the two populations are close to or lower than the minimum effective sizes conservatively recommended for management of endangered species (Lynch & Lande 1998).

Alternatively, while genetic considerations are important relative to management of endangered species, they typically affect populations on a longer time scale than ecological and other impacts that despoil or degrade habitats. Numerous examples presented in Caro & Laurenson (1994) document the situation that human-induced perturbations typically present greater challenges to most threatened and endangered species, and without question the altered habitat in the Cape Fear drainage (Howard 2003) likely presents the greatest threat to persistence of the species. Of future concern should be whether deteriorating habitat in the Cape Fear drainage synergistically impacts genetic diversity in a negative way.

Acknowledgements

We thank A. Howard, N. Jeffers, J. (Johnson) Ratcliffe, D. Laplante and D. Rabon for assistance in procuring specimens, F. Deal and D. Deal for providing accommodations during field work, T. Turner for providing advice regarding data analysis, M.A. Beaumont for providing us with the program msvar1.3 and C. Burridge for assistance in the laboratory. Comments on a draft of the manuscript were kindly provided by R. Edwards and T. Turner. Sampling was carried out under Endangered/Threatened Species Subpermittee Authorization Number SA 02-02 (US Department of the Interior). This work was supported by the US Fish and Wildlife Service (US Department of the Interior) under Agreement no. 1448-40181-00-G-119, and by the Texas Agricultural Experiment Station under Project H-6703. Views expressed in the paper are those of the authors and do not necessarily reflect views of the sponsors. The study is Contribution no. 123 of the Center for Biosystematics and Biodiversity at Texas A&M University.

© 2004 Blackwell Publishing Ltd, Molecular Ecology, 13, 2947–2958
CONSERVATION GENETICS OF N. MEKISTOCHOLAS

References


Appendix I

Summary statistics at 22 nuclear-encoded loci in two populations of Cape Fear shiners (*Notropis mekistocholas Locality*)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Confluence</th>
<th>High Falls</th>
<th>Locus</th>
<th>Confluence</th>
<th>High Falls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nme4F4.154</td>
<td>N</td>
<td>44</td>
<td>Nme4F4.222</td>
<td>N</td>
<td>44</td>
</tr>
<tr>
<td>#A</td>
<td>5</td>
<td>4</td>
<td>#A</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>AR</td>
<td>3.34</td>
<td>3.66</td>
<td>AR</td>
<td>5.90</td>
<td>5.74</td>
</tr>
<tr>
<td>H_E</td>
<td>0.546</td>
<td>0.625</td>
<td>H_E</td>
<td>0.684</td>
<td>0.761</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.882</td>
<td>0.847</td>
<td>P_HW</td>
<td>0.116</td>
<td>0.714</td>
</tr>
<tr>
<td>F_S</td>
<td>0.001</td>
<td>–0.015</td>
<td>F_S</td>
<td>0.103</td>
<td>0.023</td>
</tr>
<tr>
<td>Nme7B9.190</td>
<td>N</td>
<td>42</td>
<td>Nme18A6.158</td>
<td>N</td>
<td>44</td>
</tr>
<tr>
<td>#A</td>
<td>10</td>
<td>9</td>
<td>#A</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>AR</td>
<td>6.85</td>
<td>6.10</td>
<td>AR</td>
<td>6.48</td>
<td>5.70</td>
</tr>
<tr>
<td>H_E</td>
<td>0.840</td>
<td>0.820</td>
<td>H_E</td>
<td>0.803</td>
<td>0.776</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.777</td>
<td>0.637</td>
<td>P_HW</td>
<td>0.205</td>
<td>0.121</td>
</tr>
<tr>
<td>F_S</td>
<td>0.064</td>
<td>0.110</td>
<td>F_S</td>
<td>0.095</td>
<td>0.066</td>
</tr>
<tr>
<td>Nme18C2.178</td>
<td>N</td>
<td>44</td>
<td>Nme24B6.191</td>
<td>N</td>
<td>44</td>
</tr>
<tr>
<td>#A</td>
<td>8</td>
<td>9</td>
<td>#A</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>AR</td>
<td>4.95</td>
<td>5.44</td>
<td>AR</td>
<td>1.76</td>
<td>1.84</td>
</tr>
<tr>
<td>H_E</td>
<td>0.718</td>
<td>0.772</td>
<td>H_E</td>
<td>0.128</td>
<td>0.158</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.075</td>
<td>0.616</td>
<td>P_HW</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>F_S</td>
<td>0.114</td>
<td>0.147</td>
<td>F_S</td>
<td>–0.062</td>
<td>–0.081</td>
</tr>
<tr>
<td>Nme24B6.211</td>
<td>N</td>
<td>44</td>
<td>Nme25C8.208</td>
<td>N</td>
<td>43</td>
</tr>
<tr>
<td>#A</td>
<td>10</td>
<td>10</td>
<td>#A</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>AR</td>
<td>5.70</td>
<td>6.06</td>
<td>AR</td>
<td>6.82</td>
<td>7.20</td>
</tr>
<tr>
<td>H_E</td>
<td>0.740</td>
<td>0.663</td>
<td>H_E</td>
<td>0.812</td>
<td>0.846</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.632</td>
<td>0.281</td>
<td>P_HW</td>
<td>0.171</td>
<td>0.567</td>
</tr>
<tr>
<td>F_S</td>
<td>–0.044</td>
<td>0.080</td>
<td>F_S</td>
<td>0.112</td>
<td>–0.009</td>
</tr>
<tr>
<td>Nme26G8.96</td>
<td>N</td>
<td>44</td>
<td>Nme30D2.120</td>
<td>N</td>
<td>44</td>
</tr>
<tr>
<td>#A</td>
<td>5</td>
<td>5</td>
<td>#A</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>AR</td>
<td>3.52</td>
<td>4.21</td>
<td>AR</td>
<td>2.71</td>
<td>1.84</td>
</tr>
<tr>
<td>H_E</td>
<td>0.547</td>
<td>0.635</td>
<td>H_E</td>
<td>0.348</td>
<td>0.158</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.472</td>
<td>0.682</td>
<td>P_HW</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>F_S</td>
<td>0.169</td>
<td>–0.192</td>
<td>F_S</td>
<td>–0.112</td>
<td>–0.081</td>
</tr>
<tr>
<td>Nme18A6.0</td>
<td>N</td>
<td>43</td>
<td>Nme5B10.211</td>
<td>N</td>
<td>44</td>
</tr>
<tr>
<td>#A</td>
<td>3</td>
<td>3</td>
<td>#A</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>AR</td>
<td>2.89</td>
<td>2.64</td>
<td>AR</td>
<td>7.98</td>
<td>6.80</td>
</tr>
<tr>
<td>H_E</td>
<td>0.587</td>
<td>0.551</td>
<td>H_E</td>
<td>0.863</td>
<td>0.827</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.102</td>
<td>0.365</td>
<td>P_HW</td>
<td>0.996</td>
<td>0.892</td>
</tr>
<tr>
<td>F_S</td>
<td>0.090</td>
<td>0.203</td>
<td>F_S</td>
<td>0.026</td>
<td>–0.091</td>
</tr>
<tr>
<td>Nme27F8.230</td>
<td>N†</td>
<td>44</td>
<td>Nme12D8.125</td>
<td>N</td>
<td>44</td>
</tr>
<tr>
<td>#A</td>
<td>8</td>
<td>8</td>
<td>#A</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>AR</td>
<td>6.07</td>
<td>6.23</td>
<td>AR</td>
<td>7.98</td>
<td>6.14</td>
</tr>
<tr>
<td>H_E</td>
<td>0.757</td>
<td>0.821</td>
<td>H_E</td>
<td>0.869</td>
<td>0.781</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.623</td>
<td>0.882</td>
<td>P_HW</td>
<td>&lt; 0.001*</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>F_S</td>
<td>–0.111</td>
<td>–0.040</td>
<td>F_S</td>
<td>0.372</td>
<td>0.672</td>
</tr>
<tr>
<td>Nme15F2.174</td>
<td>N</td>
<td>44</td>
<td>Nme30F12.208</td>
<td>N</td>
<td>42</td>
</tr>
<tr>
<td>#A</td>
<td>6</td>
<td>7</td>
<td>#A</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>AR</td>
<td>4.31</td>
<td>5.20</td>
<td>AR</td>
<td>7.66</td>
<td>7.97</td>
</tr>
<tr>
<td>H_E</td>
<td>0.560</td>
<td>0.705</td>
<td>H_E</td>
<td>0.846</td>
<td>0.874</td>
</tr>
<tr>
<td>P_HW</td>
<td>–0.055</td>
<td>0.032</td>
<td>P_HW</td>
<td>0.184</td>
<td>0.256</td>
</tr>
<tr>
<td>F_S</td>
<td>0.962</td>
<td>0.547</td>
<td>F_S</td>
<td>0.017</td>
<td>0.002</td>
</tr>
<tr>
<td>Nme33B6.125</td>
<td>N</td>
<td>44</td>
<td>Nme2B10.166</td>
<td>N</td>
<td>43</td>
</tr>
<tr>
<td>#A</td>
<td>5</td>
<td>7</td>
<td>#A</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>AR</td>
<td>4.37</td>
<td>5.51</td>
<td>AR</td>
<td>4.88</td>
<td>4.76</td>
</tr>
<tr>
<td>H_E</td>
<td>0.666</td>
<td>0.738</td>
<td>H_E</td>
<td>0.727</td>
<td>0.651</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.705</td>
<td>0.132</td>
<td>P_HW</td>
<td>0.888</td>
<td>0.754</td>
</tr>
</tbody>
</table>
| F_S         | –0.058     | 0.173      | F_S         | 0.008      | –0.010     

Appendix I  Continued

<table>
<thead>
<tr>
<th>Locus</th>
<th>Confluence</th>
<th>High Falls</th>
<th>Locus</th>
<th>Confluence</th>
<th>High Falls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nme2B10.232</td>
<td>N</td>
<td>44</td>
<td></td>
<td>Nme5G4.175</td>
<td>N</td>
</tr>
<tr>
<td>#A</td>
<td>13</td>
<td>11</td>
<td>#A</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>AR</td>
<td>8.40</td>
<td>7.83</td>
<td>AR</td>
<td>7.50</td>
<td>6.11</td>
</tr>
<tr>
<td>H_E</td>
<td>0.843</td>
<td>0.871</td>
<td>H_E</td>
<td>0.812</td>
<td>0.734</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.050</td>
<td>0.705</td>
<td>P_HW</td>
<td>0.026</td>
<td>0.476</td>
</tr>
<tr>
<td>F_IS</td>
<td>0.084</td>
<td>0.076</td>
<td>F_IS</td>
<td>0.132</td>
<td>−0.030</td>
</tr>
<tr>
<td>Nme6A7.93</td>
<td>N</td>
<td>44</td>
<td></td>
<td>Nme2D5.117</td>
<td>N</td>
</tr>
<tr>
<td>#A</td>
<td>11</td>
<td>10</td>
<td>#A</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>AR</td>
<td>7.51</td>
<td>7.30</td>
<td>AR</td>
<td>6.65</td>
<td>8.07</td>
</tr>
<tr>
<td>H_E</td>
<td>0.850</td>
<td>0.848</td>
<td>H_E</td>
<td>0.824</td>
<td>0.877</td>
</tr>
<tr>
<td>P_HW</td>
<td>0.790</td>
<td>0.472</td>
<td>P_HW</td>
<td>0.747</td>
<td>0.869</td>
</tr>
<tr>
<td>F_IS</td>
<td>0.010</td>
<td>0.049</td>
<td>F_IS</td>
<td>−0.130</td>
<td>0.029</td>
</tr>
</tbody>
</table>

*Significant after Bonferroni correction.

_N_ = sample size, †_A_ = number of alleles; AR = allelic richness, _H_E_ = gene diversity, _P_HW_ = probability of conformity to Hardy–Weinberg genotypic expectations and _F_IS_ = inbreeding coefficient.