

Reconciling nuclear microsatellite and mitochondrial marker estimates of population structure: breeding population structure of Chesapeake Bay striped bass (*Morone saxatilis*)

KM Brown, GA Baltazar and MB Hamilton

Department of Biology, Georgetown University, Reiss Sciences Building 406, 37th and O Streets NW, Washington, DC 20057-1229, USA

Comparative analyses of nuclear and organelle genetic markers may help delineate evolutionarily significant units or management units, although population differentiation estimates from multiple genomes can also conflict. Striped bass (*Morone saxatilis*) are long-lived, highly migratory anadromous fish recently recovered from a severe decline in population size. Previous studies with protein, nuclear DNA and mitochondrial DNA (mtDNA) markers produced discordant results, and it remains uncertain if the multiple tributaries within Chesapeake Bay constitute distinct management units. Here, 196 young-of-the-year (YOY) striped bass were sampled from Maryland's Choptank, Potomac and Nanticoke Rivers and the north end of Chesapeake Bay in 1999 and from Virginia's Mataponi and Rappahannock Rivers in 2001. A total of 10 microsatellite loci exhibited between two and 27 alleles per locus with observed heterozygosities between 0.255 and 0.893. The 10-locus

estimate of R_{ST} among the six tributaries was -0.0065 (95% confidence interval -0.0198 to 0.0018). All R_{ST} and all but one θ estimates for pairs of populations were not significantly different from zero. Reanalysis of Chesapeake Bay striped bass mtDNA data from two previous studies estimated population differentiation between $\theta = -0.002$ and 0.160 , values generally similar to mtDNA population differentiation predicted from microsatellite R_{ST} after adjusting for reduced effective population size and uniparental inheritance in organelle genomes. Based on mtDNA differentiation, breeding sex ratios or gene flow may have been slightly male biased in some years. The results reconcile conflicting past studies based on different types of genetic markers, supporting a single Chesapeake Bay management unit encompassing a panmictic striped bass breeding population. *Heredity* (2005) **94**, 606–615. doi:10.1038/sj.hdy.6800668
Published online 13 April 2005

Keywords: management units; microsatellite; mtDNA; *Morone saxatilis*; population structure

Introduction

Conservation genetics seeks to quantify genetic differentiation among populations using molecular markers that serve as a surrogate for a wider range of biological factors such as some degree of reproductive or demographic independence, historical independence of populations due to limited gene flow or adaptive divergence due to unique ecological conditions. Genetic differentiation among populations is therefore a key condition in the recognition of evolutionarily significant units (ESUs) and management units (MUs) in protected and managed populations (Ryder, 1986; see reviews in Moritz, 1994; Moritz *et al.*, 1995; Crandall *et al.*, 2000; Fraser and Bernatchez, 2001 and references therein). Although a great deal of attention has been focused on alternative conceptual definitions of ESUs and MUs, the estimates of population differentiation (or lack thereof) that are prerequisites to employing these concepts may not be

unambiguous. Conservation genetics now employs a wide range of possible technical approaches and genetic marker types (eg Parker *et al.*, 1998; Wayne and Morin, 2004). An increasing number of studies use comparative analyses that capitalize on the different effective population sizes and mutation rates of nuclear and uniparentally inherited organelle genetic markers to estimate population differentiation as well as to test hypotheses related to sex-specific differences in gene flow or sex ratio bias (see Hamilton and Miller, 2002) in the process of delineating ESUs and MUs. Examples of analyses comparing nuclear and organelle genetic markers to estimate population differentiation and infer components of gene flow include studies of wild dogs (Girman *et al.*, 2001), blue marlin (Buonaccorsi *et al.*, 2001), red drum (Gold and Turner, 2002), frogs (Crawford, 2003), elephants (Nyakaana and Arctander, 1999), prairie-chickens (Johnson *et al.*, 2003) and a temperate forest tree (Oddou-Muratorio *et al.*, 2001).

Mid-Atlantic populations of anadromous striped bass (*Morone saxatilis*) are a case where employment of different classes of molecular markers has led to conflicting estimates of population differentiation and thereby difficulty in defining management units. Striped bass populations experienced a precipitous decline after 1973 that continued through the 1980s, forcing drastic

changes in population management and a surge of research to better estimate genetic and demographic population parameters, with striped bass populations having recently been declared recovered (Richards and Rago, 1999; ASMFC, 2003). Studies using both mitochondrial (Wirgin *et al*, 1990, 1997) and nuclear (Bielawski and Pumo, 1997) DNA markers clearly support marked genetic differentiation among breeding populations in the Hudson River, Delaware River, Chesapeake Bay and Roanoke River. However, among the multiple tributaries within Chesapeake Bay, the estuary system that contributes as much as 90% of recruitment to the coastal population (Waldman *et al*, 1997), striped bass management units have been difficult to define biologically.

Studies using protein polymorphisms, nuclear DNA and organelle DNA to estimate differentiation among striped bass tributary breeding populations in Chesapeake Bay have produced conflicting results. Morgan *et al* (1973) first used serum proteins to show differentiation among striped bass sampled in Maryland tributaries. A later study of fish from Maryland tributaries and the Rappahannock River based on serum proteins as well as the G-3-PDH protein showed no evidence for genetic populations (Sidell *et al*, 1980). Chapman (1989, 1990) showed mitochondrial DNA (mtDNA) restriction endonuclease fragment length polymorphism (RFLPs) frequency differences among sampling locations within Chesapeake Bay, which he attributed to stronger homing to natal tributaries among females than among males. Wirgin *et al* (1990) also examined mtDNA RFLPs among mid-Atlantic striped bass and found evidence that upper and lower Chesapeake Bay tributaries exhibited distinct length haplotypes. Most recently, Laughlin and Turner (1996) showed very little genetic differentiation for multilocus variable number of tandem repeat (VNTR) markers among young-of-year (hereafter YOY) fish from the James, Rappahannock and York rivers in the Virginia portion of Chesapeake Bay. The mtDNA studies are widely interpreted as evidence of modest genetic differentiation among Chesapeake Bay tributaries. Since highly polymorphic nuclear genetic markers have been difficult to develop for *Morone* species (Han *et al*, 2000; Han and Ely, 2002), one hypothesis to explain a lack of population differentiation in nuclear DNA markers is that modest genetic differentiation exists among tributary breeding populations but it is difficult to demonstrate due to lack of genetic marker allelic diversity. An alternative hypothesis is that nuclear and organelle marker estimates are actually consistent after compensating for reduced effective population size and uniparental inheritance in organelle genomes.

Here we present a multilocus nuclear genetic marker estimate of population differentiation among striped bass breeding populations in six Chesapeake Bay tributaries utilizing 10 recently developed moderately to highly polymorphic microsatellite loci. YOY or age class zero fish were used exclusively since they have limited capacity for dispersal and therefore were most likely the offspring of spawning in the portion of the Bay where they were sampled. Estimating genetic differentiation with YOY fish controls for age structure within the adult population since there is no year-class heterogeneity within the sampled individuals. Further, YOY populations represent a weighted average of allele frequencies of those fish that actually bred among all

the fish present on the spawning grounds and should exhibit spatial genetic patterns dictated by mating patterns in the parental populations. Utilizing YOY fish and highly polymorphic genetic markers should increase the resolution to detect modest population structure among Chesapeake Bay striped bass breeding populations if it exists. Based on nuclear microsatellite data and a reanalysis of mtDNA data for striped bass, we compare expected levels of population differentiation for nuclear and organelle markers under an island model of population structure (Hamilton and Miller, 2002). Our goal is to reconcile the conflict between nuclear and organelle estimates of population structure in striped bass and thereby estimate the boundaries of biological management units for striped bass within Chesapeake Bay. We also highlight ways in which expected levels of population differentiation may differ between nuclear and organelle loci as an aid to comparative genetic marker analyses in conservation genetics.

Materials and methods

Species and sample collection

Striped bass (*Morone saxatilis*) are a long-lived, marine fish found along the North American Atlantic coast from Florida to Nova Scotia. Mid-Atlantic populations between North Carolina and Maine are anadromous and migratory, while smaller, often disjunct populations found south of Cape Hatteras (eg Bulak *et al*, 1997), along the eastern Gulf of Mexico (eg Wirgin *et al*, 1991) and between the southern tip of Nova Scotia and the St Lawrence River (Rulifson and Dadswell, 1995) are thought to be largely estuarine to riverine. Mid-Atlantic stock striped bass are recruited from breeding populations in multiple tributaries of Chesapeake Bay, the Delaware River and the Hudson River (Waldman *et al*, 1997) and are the basis of a large commercial and recreational fishery. Mid-Atlantic striped bass thus constitute a classical mixed stock fishery where adults are harvested as one stock that is potentially composed of fish recruited from geographically discrete, potentially genetically distinct breeding populations (Utter and Ryman, 1993).

YOY striped bass were collected from Maryland and Virginia tributaries of the Chesapeake Bay (Figure 1, Table 1). Samples from 18 permanent sampling stations within the Maryland region of the Chesapeake Bay were collected in August 1999 by Maryland Department of Natural Resources, Striped Bass Stock Assessment Project biologists (see MD DNR 2002). This annual juvenile survey documents year-class success for YOY striped bass, using both vessel-dragged and manually dragged seine nets. Samples from the Virginia portion of the Bay were collected from three sites in September 2001 using a manually dragged beach seine. Samples from both regions were collected at various times during the day and at varying depths. Table 1 gives sampling locations and sample sizes at each location and within each tributary. All individuals were identified as YOY by length (between 45 and 130 mm). In the field, individuals were placed in Ziploc bags on ice. Once in the laboratory, fish were re-sorted to verify identification, and individual fish were placed in whirl-pak style bags labeled

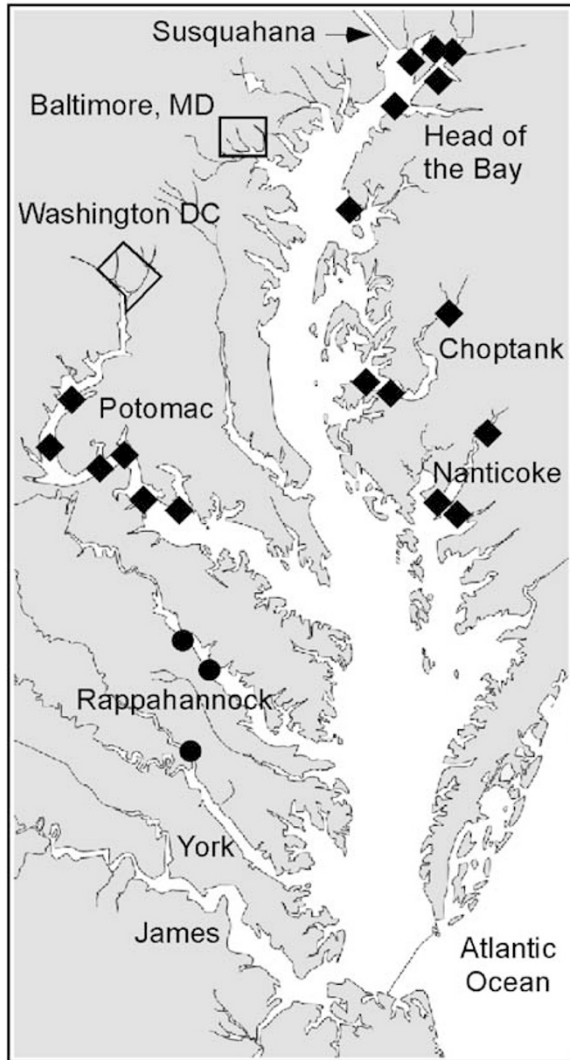


Figure 1 A map of Chesapeake Bay and its tributaries. YOY sampling sites are shown in Maryland (◆) and in Virginia (●). The Maryland locations are long-term sites used by the Maryland Department of Natural Resources (MD DNR). Washington, DC and Baltimore, MD are shown as reference points.

with collection site, date and an individual identification number and stored at -20°C .

Genomic DNA from sampled YOY fish was extracted using either the DNeasy kit (Qiagen, Vaencia, CA, USA) or Puregene DNA Isolation Kit (Gentra, Minneapolis, MN, USA) according to the manufacturer's protocols. For both extraction methods, 5–10 mg of striped bass muscle tissue was incubated overnight at 55°C with 1.5 μl proteinase K (20 mg/ml) and the following day incubated for 60 min with 1.5 μl RNase A (100 mg/ml). Genomic DNA extractions were visualized on 1% agarose, $1 \times$ TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) gels stained with ethidium bromide and repeated if the sample failed to yield a visible, high molecular weight band.

Microsatellite loci and genotype collection

A total of 11 microsatellite loci were used to genotype the YOY fish: the nine loci described by Brown *et al*

(2003), the SB111 locus described by Roy *et al* (2000) and the SB06 locus described by Han *et al* (2000). A new reverse primer for SB111 was designed and employed to improve amplification (5' ATC TGA CAG CTA ACG TCC CC 3'). Primer sets for each locus were ordered HPLC-purified with a forward primer label of 6-FAM, HEX (Operon Technologies, Alameda, CA, USA), NED, PET or VIC (Applied Biosystems, Foster City, CA, USA).

Following conditions given by Brown *et al* (2003), loci were amplified by PCR in 20 μl reactions containing 2–3 μl of DNA template (DNA concentrations not determined), 2 μl of $10 \times$ Thermopol buffer (containing 20 mM MgSO_4), 0.2 mM of each dNTP and 0.4 μM of each primer. General thermal cycling profiles were 5 min at 96°C followed by cycles of 96°C for 45 s, annealing temperature of 58°C for 60 s and 72°C extension period. For increased specificity, SB06 and SB111 PCR amplification conditions were modified from those originally reported. SB111 reactions used 0.4 μM of each primer and 1.25 M betaine (Aldrich, cat. num. 21,906-1, Milwaukee, WI, USA) and an extension time of 30 s with a total of 29 amplification cycles. SB06 reactions used 0.2 μM of each primer, a total MgSO_4 concentration of 4 mM, an extension time of 15 s and a total of 30 amplification cycles. If amplification was weak for individual samples, PCR modifications included increasing MgSO_4 concentration (up to 6 mM) and increasing the number of amplification cycles. When available, PCR product of the insert from the original microsatellite plasmid clones was diluted 1000:1 to 10000:1 and used as positive control templates to verify PCR product molecular weight. All reactions were visualized on 2% agarose and $1 \times$ TBE gels stained with ethidium bromide to verify the success of amplification in the appropriate size range.

To score genotypes, PCR products from three or four loci amplified independently from the same individual were diluted between 10- and 200-fold in one aliquot of sterile water. A measure of 1 μl of this dilution was added to 10 μl of formamide containing 0.1 μl of GeneScan ROX 400 or LIZ 500 molecular weight marker (Applied Biosystems, Foster City, CA, USA). The sample was heat denatured at 96°C for 2 min, placed on ice for at least 5 min and then electrophoresed on an ABI 3100 capillary sequencer (Applied Biosystems, Foster City, CA, USA). DNA fragment molecular weights were estimated using GeneScan 3.7 and the local Southern sizing method. PCR products too weak to be accurately sized were either rerun with less dilution or reamplified to produce a stronger PCR band. PCR products so intense as to be 'off-scale' were further diluted and electrophoresed again. In all fragment-sizing runs, the sizing standard curve was checked to verify a high coefficient of determination (r^2) and detection of all sizing DNA fragments. The number of microsatellite repeats for each fragment was determined by comparing the size (in base pairs) of DNA fragments with the size of a fragment from the same locus with a known number of repeats determined by DNA sequencing. The size difference divided by the length of the microsatellite repeat at that locus determined the number of repeats for a fragment. DNA fragments that deviated substantially from expected allele sizes were subjected to repeated electrophoresis or PCR or both to verify fragment sizes.

Table 1 Sampling location, sampling dates and number of YOY fish genotyped from each tributary or region of Chesapeake Bay

Tributary	Site name	Location	Date	N
Choptank	Castle Haven (CH)	38°34' N, 76°12' W	8/13/99	50
	Mouth of Tuckahoe (MOT)	38°47' N, 75°54' W	8/13/99	24
	Hambrook's Bar (HB)	38°33' N, 76°04' W	8/13/99	22
				4
Head of the Bay	Howell Point (HP)	39°21' N, 76°06' W	8/16/99	31
	Tolchester (T)	39°11' N, 76°15' W	8/16/99	11
	Elk Neck (E)	39°28' N, 75°56' W	8/17/99	5
	Hyland Point (HY)	39°31' N, 75°54' W	8/17/99	4
	Parlor Point (PP)	39°26' N, 75°52' W	8/17/99	5
	Carpenter's Point (CP)	39°30' N, 76°00' W	8/18/99	1
				5
Potomac	Blossom Point (BP)	38°25' N, 77°07' W	8/11/99	50
	Indianhead (IH)	38°34' N, 77°11' W	9/11/99	12
	Hallowing Point (HA)	38°38' N, 77°10' W	8/11/99	11
	Liverpool Point (LP)	38°27' N, 77°18' W	8/11/99	7
	Morgantown (M)	38°20' N, 77°01' W	8/12/99	8
	Rock Point (RP)	38°16' N, 77°53' W	8/12/99	7
Nanticoke				5
	Tyaskin Beach (TB)	38°20' N, 75°53' W	8/10/99	26
	Sharptown (S)	38°32' N, 75°44' W	8/10/99	18
Mataponi (upper York)	Lewis Landing (LL)	38°24' N, 75°53' W	8/10/99	3
				5
				7
Rappahannock	Melrose Landing (ML)	37°38' N, 76°51' W	9/8/01	7
				7
Carter's Wharf (CW)				32
	Jones Creek (JC)	38°04' N, 76°55' W	9/8/01	23
		38°01' N, 76°54' W	9/8/01	9
Total (all sites)				196

Statistical methods

We used Genepop (Raymond and Rousset, 1995; <http://wbiomed.curtin.edu.au/genepop/>) to compare observed genotype frequencies to Hardy–Weinberg expected frequencies in both the total sample of individuals and within each of the six separate populations. Additionally, to determine if each of the microsatellite loci represents independent linkage units, we employed the contingency table test in Genepop to determine if genotypes at one locus were independent of genotypes at another locus for all pairs of the 10 microsatellite loci.

We estimated population structure using two distinct methods, each with different assumptions. First, we employed R_{ST} (Slatkin, 1995) to estimate population differentiation under the assumption of a stepwise mutation model to approximate the mutation process at microsatellite loci. R_{ST} utilizes the variances in allelic state (number of repeats) within and among populations to estimate population differentiation. Under the stepwise mutation model (see review by Estoup *et al*, 2002), alleles closer in state are more likely to be recently identical by descent (IBD) but diverged in state due to mutation. Therefore, alleles more similar in state contribute less to R_{ST} than do alleles more divergent in state given identical frequency differences among populations. The estimation method for R_{ST} followed Goodman (1997) and employed standardized allele sizes using code written in Matlab (R13, v. 6.5). A 95% confidence interval for the multilocus estimate of R_{ST} was estimated by 5000 bootstrap resampling iterations over populations with replacement (Efron and Tibshirani, 1993). This bootstrap procedure estimates the sampling variance in R_{ST} from randomly drawn subsamples of the sampled

populations, which is equivalent to treating population as a random effect when carrying out a hypothesis test for population differentiation. In addition, both R_{ST} and θ , an infinite alleles model estimator of genetic differentiation (Weir, 1996), were estimated for all pairs of tributary populations. Using Matlab code, estimates of R_{ST} for pairs of populations were compared to a null distribution generated by random permutation where all individuals from the two observed populations were randomly assigned without replacement to two populations of the original sizes. R_{ST} was estimated from each of 5000 such random permutations to create a 95% confidence interval. Estimates of θ for pairs of populations with 95% confidence intervals obtained by bootstrapping across loci for 5000 iterations were obtained with the program GDA (Lewis and Zaykin, 2001). Both methods to estimate 95% confidence intervals treat population as a fixed effect when carrying out a hypothesis test for population differentiation.

Second, we used the Bayesian genotype clustering program Structure v2.0 (Pritchard *et al*, 2000; <http://pritch.bsd.uchicago.edu/software.html>) to determine the most likely number of genetic populations given the observed genotype data. Assuming Hardy–Weinberg equilibrium within populations, the model-based clustering method employed by Structure v2.0 assigns individuals to populations probabilistically based on their multilocus genotypes and thereby estimates the posterior probability for a given number of genetic populations (K) (Pritchard *et al*, 2000). The most likely value of K is the one that maximizes the log-likelihood of obtaining the observed sample of multilocus genotypes. Following the application of this program by Rosenberg

et al (2001), we used 50 000 steps with a burn-in of 5000 for each K tested and ran the program 20 times for each K value. All possible values of K (1–6 with 6 being the total number of sampled tributaries) were considered equally likely, equivalent to a uniform prior distribution of the number of populations.

For comparison to nuclear microsatellite population differentiation results, we conducted a reanalysis of striped bass mitochondrial haplotype frequency data for multiple Chesapeake Bay sampling locations originally presented by Chapman (1989) and Wirgin *et al* (1990, 1997). These studies presented haplotype frequency data but not estimates of population genetic differentiation as fixation indices. Using these data we estimated mtDNA fixation among populations with analysis of variance (θ ; see Weir, 1996) using code written in Matlab as well as with the program Arlequin (Schneider *et al*, 2000). From Wirgin *et al* (1990), we used the frequency of five RFLP haplotypes in their Table 3 for the James, Rappahannock, Choptank and Upper Chesapeake Bay. From Wirgin *et al* (1997), we used the frequency of four 'mtDNA major length variants' and seven 'mtDNA composite major length variants and *TaqI* variants' from Table 5 for the Choptank, Potomac, Rappahannock and Upper Chesapeake Bay. From Chapman (1989), we obtained the frequency of five RFLP haplotypes from his Table 1 for the Potomac, Choptank and Upper Bay in 1984, 1986 and 1987, excluding data from Hart-Miller Island in 1985. As for R_{ST} , 95% confidence intervals for θ were estimated by 5000 bootstrap resampling iterations over populations with replacement.

The multilocus genotype data and the Matlab code used to estimate R_{ST} and haploid θ are available at <http://bioserver.georgetown.edu/faculty/hamilton> under the 'downloads' link.

Results

Sampling locations and sample sizes for each tributary and at each sampling location within tributaries are given in Table 1. Sample sizes ranged from 50 individuals in the Choptank and Potomac to seven individuals in the Mataponi, an upstream tributary of the York River. The total sample size from all six tributaries was 196 individuals.

Preliminary screening of several of the most polymorphic loci (SB731, SB91, SB113) reported by Roy *et al* (2000) showed little or no polymorphism in Chesapeake Bay fish, so these loci were not tested further. In our sample of 196 fish, the locus AT150-2#10 reported by Brown *et al* (2003) was found to be monomorphic and was excluded from further analyses. For the 10 polymorphic loci, we observed between two and 27 alleles per locus with heterozygosities ranging from 0.255 to 0.893 (Table 2). These 10 loci separate approximately into three groups based on observed polymorphism (Table 2). Four loci (AG25-1#1, AC25-6#12, AT150-2#21, SB06) exhibited high levels of polymorphism with 15–27 alleles per locus. Moderate levels of polymorphism were observed at three loci (AT150-2#4, AT150-2#12, AC25-6#10) with 4–7 alleles per locus. Three loci (AT150-2#11, AC 25-1#2, SB111) exhibited low levels of polymorphism with only two or three alleles per locus. Figure 2 shows histograms of allele frequency for each locus.

Three loci had genotype frequencies that deviated from Hardy–Weinberg expectations (Table 2). However, in tests for Hardy–Weinberg equilibrium within individual tributary populations, only AT150-2#21 deviated from Hardy–Weinberg expectations in more than a single tributary (Table 2). In 43 of 45 contingency table tests comparing pairs of loci, genotype frequencies were independent within populations. Two exceptions were nonindependence of genotype frequencies between loci AC25-1#2 and AC25-6#12 within the Nanticoke ($P < 0.001$) and between loci AT150-2#4 and AC25-6#10 within the Choptank ($P < 0.001$). For the latter two loci, the Head of the Bay ($P < 0.07$), Nanticoke ($P < 0.07$) and Rappahannock ($P < 0.08$) approached nonindependence of genotype frequencies.

The genotype data showed instances of alleles unique to a single population, or private alleles. Across all 10 loci, each population had at least one (Mataponi) and as many as 10 (Choptank and Potomac) private alleles. Seven of the 10 loci demonstrated at least one private allele and the number of private alleles at each locus ranged from zero (SB111, AT150-2#11, AC25-6#10) to 10 (AC25-6#12). The number of private alleles per locus was correlated with observed heterozygosity at that locus ($r = 0.805$, $P < 0.05$) and the number of private alleles per population was correlated with population sample size ($r = 0.94$, $P < 0.05$).

Table 2 Locus name, repeat motif, observed size range, number of alleles (k), observed heterozygosity (H_o) and Hardy–Weinberg test results (HW) for the 10 microsatellite loci used to estimate population structure

Locus	Repeat	Size range (bp)	k	H_E	H_o^a	HW ^b
AG25-1#1	CTTT	154–308	26	0.905	0.893	—
AT150-2#11	GT	208–212	2	0.320	0.255*	C
AC25-1#2	AC	85–97	3	0.454	0.526	—
AT150-2#4	GT	147–161	7	0.558	0.439*	H
AC25-6#12	TG	161–209	18	0.681	0.689	—
AT150-2#12	GT	117–127	5	0.477	0.490	—
AT150-2#21	GT	168–196	15	0.788	0.633*	C, N, R
AC25-6#10	AT/GT	218–224	4	0.396	0.398	—
SB111	TGTT	101–105	2	0.411	0.459	—
SB06	GT	173–243	27	0.803	0.724	—

^aObserved heterozygosity of the total sample of 196 individuals.

^bIndicates the specific tributary populations that deviate from Hardy–Weinberg (C: Choptank; H: Head of the Bay; N: Nanticoke; R: Rappahannock) for those loci that deviate from Hardy–Weinberg expectations at the total population level.

*Observed heterozygosity significantly different from Hardy–Weinberg expectations ($P > 0.05$).

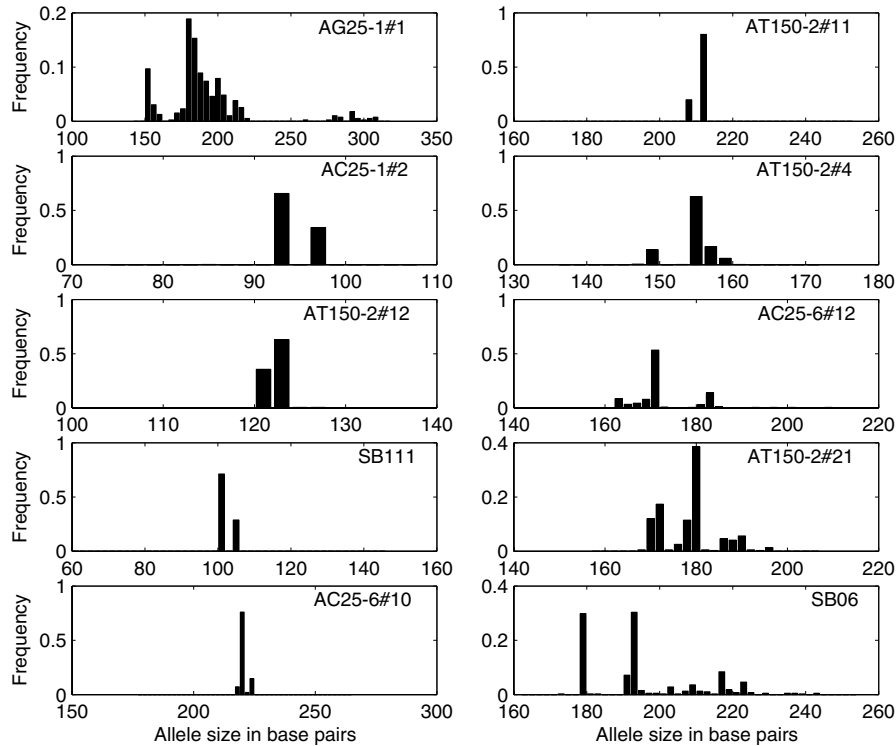


Figure 2 Histograms of allele sizes in base pairs for each of the 10 microsatellite loci.

Table 3 Genetic differentiation between pairs of populations estimated with R_{ST} (above diagonal) assuming a stepwise mutation model and θ (below diagonal) assuming an infinite alleles mutation model

	C	H	P	N	R	M
Choptank	—	-0.0004	-0.0038	-0.0077	0.0107	-0.0034
Head of the Bay	0.0048	—	-0.0012	-0.0046	0.0035	-0.0097
Potomac	0.0161	0.0313	—	-0.0050	0.0032	0.0082
Nanticoke	0.0049	0.0076	0.0197	—	0.0148	-0.0008
Rappahanock	0.0051	0.0036	0.0275 ^a	-0.0004	—	-0.0036
Mataponi	0.0048	0.0111	0.0298	0.0025	-0.0036	—

C: Choptank; H: Head of the Bay; P: Potomac; N: Nanticoke; R: Rappahanock; M: Mataponi (upper York).

^aThe 95% confidence interval estimated with a randomization test (see text) does not include zero.

Table 4 The posterior probability distribution for the number of populations (K) given the observed genotype frequencies (X). Each possible value of K was assumed to be equally likely (a uniform prior distribution of K)

K	$\ln P(K X)$	$P(K X)$
1	-5048	~1.0
2	-5170	<10 ⁻⁵⁰
3	-5518	<10 ⁻¹⁰⁰
4	-5670	<10 ⁻¹⁰⁰
5	-6422	<10 ⁻¹⁰⁰
6	-6636	<10 ⁻¹⁰⁰

The 10-locus estimate of R_{ST} among the six tributaries was -0.0065 with a 95% confidence interval of -0.0198 to 0.0018. This point estimate is not significantly different from zero, indicating no population subdivision. Using only one locus from each of the pairs of loci which

exhibited nonindependence of genotype frequencies, the four possible eight-locus estimates of R_{ST} were qualitatively identical to the 10 locus estimate. All R_{ST} estimates and all but one θ estimate for pairs of populations fell within their respective 95% confidence intervals determined by random permutation (Table 3). Given that Table 3 contains 30 nonindependent pairwise comparisons, a single rejection of the null hypothesis is probably not biologically meaningful. At a significance level of 0.05, we expect 1 of 20 tests to reject the null hypothesis even when it is true (see Rice, 1989). Bayesian genotype clustering analysis estimated a single genetic population to be most likely, with all other possible values of K having vanishingly small likelihoods (Table 4).

The mtDNA data of Wirgin *et al* (1990) produced an estimate of $\theta = -0.002$ with a 95% confidence interval of -0.081 to 0.039. Data from Wirgin *et al* (1997) produced estimates of $\theta = -0.002$ (95% CI of -0.018 to 0.006) for the major length variants and $\theta = 0.042$ (95% CI of -0.018 to

0.049) for the combination of length and *TaqI* haplotypes. The data of Chapman (1989) produced estimates of mtDNA population differentiation in 3 years: 1984: $\theta = 0.138$ (95% CI 0.015–0.273); 1986: $\theta = 0.016$ (95% CI –0.041 to 0.019); 1987: $\theta = 0.160$ (95% CI 0.043–0.212). Estimates of mtDNA differentiation from Arlequin were identical to within rounding error.

Discussion

All methods to estimate genetic differentiation from these microsatellite genotype data lead to the conclusion that there is little genetic differentiation among striped bass in different tributaries within Chesapeake Bay. The genetic markers employed here demonstrate high allelic diversity and are among the most polymorphic reported to date for striped bass. This allelic diversity taken together with the generally observed high mutation rates of microsatellite loci (Ellegren, 2000) suggests that this study would be able to detect relatively modest population differentiation if it existed among striped bass populations. However, since our sample size was small in the Mataponi, this study would not be able to detect modest allele frequency differentiation in that population. We did not consider the observation of private alleles within the populations to be evidence of population differentiation since the frequency of private alleles was correlated with observed heterozygosity at each locus as well as with population sample size. These correlations are consistent with the frequency of private alleles being a product of the strongly unequal sample sizes among populations and heterogeneous mutation rates among the microsatellite loci rather than with limited gene flow.

We do not interpret the nonindependence of genotype frequencies at two pairs of loci as evidence of linkage disequilibrium between those loci for several reasons. First, physical linkage between loci should produce linkage disequilibrium within all populations given panmixia, whereas we only observed nonindependence in one population for loci AC25-1#2 and AC25-6#12. The comparisons between AT150-2#4 and AC25-6#10 showed high frequency of a combination of the same two homozygous genotypes in all populations, consistent with low frequency null alleles for those genotypes that produced a deficit of heterozygotes. Lastly, all significant deviations from Hardy–Weinberg genotype frequencies were due to deficits of heterozygotes, again consistent with null alleles.

The method of fragment analysis used here to determine microsatellite genotypes may not detect all genetic variation present at the loci sampled. Microsatellite electromorph size homoplasy is a condition where PCR fragments identical in electrophoretic mobility are not identical in DNA sequence and are therefore not IBD (reviewed by Estoup *et al*, 2002). Electromorph size homoplasy has the potential to bias estimates of population structure because genetic variation within identical electromorphs can be obscured. For all 11 microsatellite loci employed here, no instances of electromorph homoplasy were observed in a study that determined the DNA sequence of a sample of electromorphs at each locus (Adams *et al*, 2004). Thus, electromorph size homoplasy is infrequent at these loci in these populations and contributes no measurable bias

to estimated population structure. The use of R_{ST} , which assumes a stepwise mutation model, to estimate population genetic differentiation accounts for the distinct phenomenon of size homoplasy – where alleles identical in state are not IBD and *vice versa* – which is expected to be relatively frequent due to high mutation rates for microsatellites (Ellegren, 2000). In addition, an analysis of genotyping error for the 10 loci employed here (as recommended by Bonin *et al*, 2004) suggests that such errors would be unlikely to bias the estimate of R_{ST} by more than about 1% (Adams *et al*, 2004).

Differences between our results and previous studies may have a basis in the sampling methodology and striped bass life history. Specifically, the complex life history of striped bass, which includes overlapping generations, fecundity which varies with age, as well as the possibility of differential dispersal of males and females complicates estimates of breeding population structure. Fish sampled from spawning sites are very likely to be males, where males can outnumber females by 100 to 1 (Chapman, 1990; H Hornick, personal communication), but male and female dispersal patterns are likely heterogeneous. The tagging study of Dorazio *et al* (1994) indicated that smaller striped bass were less likely to migrate outside Chesapeake Bay and there was a nonsignificant trend that females were about twice as likely to disperse as males. Therefore, samples consisting predominantly of young male striped bass could represent population structure of individuals during a limited part of the life history of one sex. The ability of striped bass to disperse (Waldman *et al*, 1990; Dorazio *et al*, 1994) as well as female fecundity (Berlinsky *et al*, 1995) increases as individuals grow larger. Larger female fish produce larger eggs and the resulting larvae are also larger and have a greater chance of survival (Zastrow *et al*, 1989; Monteleone and Houde, 1990). Therefore, larger fish are both more likely to disperse greater distances and also more likely to make the greatest reproductive contribution to a given age class. However, these larger fish are usually the least frequent in samples of adults that have been used for genetic differentiation estimates. Most genetic studies of striped bass have not controlled for these aspects of life history variation since they pooled samples from a range of age classes that were sampled within Chesapeake Bay at various times of the year.

Our sample was composed entirely of YOY individuals so that estimates of population differentiation are therefore less likely to be confounded by age structure or strong sex ratio bias among the sampled fish. The only other study of striped bass genetic population structure in Chesapeake Bay that employed YOY individuals was that of Laughlin and Turner (1996). Our sampling method is a potential limitation in that it combined YOY fish sampled in two different years (1999 from Maryland and 2001 from Virginia), confounding geographic location and year of sample. However, there is no evidence that allele frequencies are heterogeneous over years as evidenced by the lack of significant differentiation over all populations or when pairs of Virginia and Maryland populations are compared. It is possible that population differentiation varies over years in striped bass due to factors such as stochastic demographic variation, variation in recruitment success and genetic drift. The existence of annual variation in

population differentiation would help to explain why past studies of striped bass have found evidence both for and against genetic divergence of striped bass populations within Chesapeake Bay. Our samples were taken in years when the geometric mean among all survey sites of the juvenile index (mean YOY striped bass per net haul) was 5.3 (1999) and 12.6 (2001) in Maryland waters (MD DNR 2002). Variation in juvenile recruitment levels among populations as well as variation in mean recruitment levels may influence the genetic structure of populations. Sampling genetic data from additional YOY year classes would provide a more complete picture of population structure over time.

Our results are most directly comparable to those of Laughlin and Turner (1996), who examined multilocus VNTR markers among YOY fish sampled in 1990 from the James, Rappahannock and York rivers in the Virginia portion of Chesapeake Bay. Their data suggested little population genetic differentiation based on a sample of YOY fish and highly variable multilocus DNA 'fingerprint' genetic markers. Although highly variable, VNTR markers can produce biased estimates of population structure since differences in band sizes are not equivalent to Mendelian alleles and using band sharing among VNTR profiles to estimate heterozygosity leads to downward bias in estimates of fixation indices (see Hamilton, 1997). However, the results presented here (based on microsatellite loci that exhibit strictly Mendelian alleles) are in qualitative agreement that very little genetic variation is found among populations within Chesapeake Bay.

The mtDNA RFLP marker studies of Chapman (1989, 1990) and Wirgin *et al* (1990, 1997) have been interpreted as evidence of genetic differentiation among Chesapeake Bay tributaries. Because organelle genomes are uniparentally inherited and haploid, their effective population size is smaller than those of nuclear genomes. Therefore, population differentiation for organelle genetic markers is expected to be higher at equilibrium than population differentiation for nuclear genetic markers, even if female and male components of gene flow are equal. In fact, equilibrium mtDNA population differentiation is expected to approach a level four times greater than nuclear population differentiation under symmetric rates of male and female gene flow and an equal sex ratio as $N_e m$ (the product of the effective population size and the rate of gene flow) approaches large values in an island model (Hamilton and Miller, 2002). If striped bass females are more faithful to breeding tributaries than males, for example, then mtDNA population differentiation should be even greater than expected under the hypothesis of equal female and male gene flow because the effective size within a local population would decrease for a matrilineal genetic marker. Based on the upper 95% confidence limit estimated here for nuclear microsatellite loci (0.0018), we expect mtDNA subdivision approaching about $\theta = 0.0072$ if male and female gene flow is identical and the sex ratio is equal.

Reanalysis of mtDNA data from Chapman (1989) produced estimates of θ significantly greater than 0.0072 in 1984 and 1987, while in 1986 the 95% confidence interval included 0.0072. In both Wirgin *et al* studies (1990, 1997), estimates of θ were not significantly different from zero and thus do not support mtDNA population structure among Chesapeake Bay tributaries.

Thus, some past estimates of mtDNA population differentiation suggest very modest deviation from levels of mtDNA population differentiation predicted from the nuclear population differentiation estimated here. The deviation is consistent with Chapman's (1989, 1990) hypothesis of 'asymmetric homing' whereby females more faithfully return to natal tributaries to breed while males disperse at random. The pattern could also be explained by a slight breeding sex bias toward fewer females. A stronger test of this hypothesis would require estimates of both nuclear and mtDNA population subdivision from a single sample of YOY fish to account for possible annual heterogeneity in breeding, population differentiation and sampling variance.

This study serves as an example of how comparative analyses of nuclear and organelle genetic markers can help to place population structure estimates into a broader context, improving the delineation of management units. Due to being haploid and to frequent uniparental inheritance, organelle DNA markers are expected to exhibit higher levels of population differentiation than nuclear DNA markers at approximate drift-migration equilibrium (Hamilton and Miller, 2002). It is therefore possible that an mtDNA population structure estimate would be significantly greater than zero while a nuclear population structure estimate would not differ significantly from zero. In marine species egg, larval and adult dispersal ability is often high and physical barriers to migration are frequently limited or absent. This leads to the expectation that genetic differentiation among populations will often be low (Waples, 1998), making the comparison of expected levels of differentiation for different types of genetic markers particularly important to recognize. A conclusion of significantly differentiated populations and thus the existence of separate management units (see Waples, 1998) may require some qualification if the data are from organelle markers and the level of differentiation is modest. In such a case, nuclear markers are unlikely to show evidence of population differentiation. Comparative genetic marker analyses can help to establish biological context for estimates of population structure as well as provide unique tools to understand and estimate the action of mechanisms such as the components of gene flow and breeding sex ratio that ultimately cause population structure and are critical parameters to monitor in managed populations.

Acknowledgements

Financial support was provided by Georgetown University, the Washington Biologists Field Club, the Georgetown-Howard Hughes Medical Institute Undergraduate Research Scholars program and a Zukowski Undergraduate Research Scholarship to KMB. Fish samples were kindly provided by H Hornick and the Maryland Department of Natural Resources, Resource Management Division, Striped Bass Stock Assessment Team. We thank C Lund for technical assistance, H Hornick for advice and support and anonymous reviewers for helpful comments. The ABI 3100 instrument was supported by an award from the National Science Foundation (DBI-0100061) and Georgetown University.

References

- Adams RI, Brown KM, Hamilton MB (2004). The impact of microsatellite electromorph size homoplasy on multilocus population structure estimates in a tropical tree (*Corythophora alta*) and an anadromous fish (*Morone saxatilis*). *Mol Ecol* **13**: 2579–2588.
- ASMFC (Atlantic States Marine Fisheries Commission) (2003). *Striped bass technical committee advisory and summary reports on the status of the Atlantic striped bass*. Atlantic States Marine Fisheries Commission, Washington, DC.
- Berlinsky DL, Fabrizio MC, O'Brien JF, Specker JL (1995). Age-at-maturity estimates for Atlantic coast female striped bass. *Trans Am Fish Soc* **124**: 207–215.
- Bielawski JP, Pumo DE (1997). Randomly amplified polymorphic DNA (RAPD) analysis of Atlantic coast striped bass. *Heredity* **78**: 32–40.
- Bonin A, Bellemain E, Bronken P, Pompanon F, Brochmann C, Taberlet P (2004). How to track and assess genotyping errors in population genetics studies. *Mol Ecol* **13**: 3261–3273.
- Brown KM, Baltazar GA, Weinstein BN, Hamilton MB (2003). Isolation and characterization of nuclear microsatellite loci in the anadromous marine fish *Morone saxatilis*. *Mol Ecol Notes* **3**: 414–416.
- Buonaccorsi VP, McDowell JR, Graves JE (2001). Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Mol Ecol* **10**: 1179–1196.
- Bulak JS, Crane JS, Secor DH, Dean JM (1997). Recruitment Dynamics of Striped Bass in the Santee-Cooper system, South Carolina. *Trans Am Fish Soc* **126**: 133–143.
- Chapman RW (1989). Spatial and temporal variation of mitochondrial haplotype frequencies in the striped bass 1982 year class. *Copeia* **1989**: 344–348.
- Chapman RW (1990). Mitochondrial DNA analysis of striped bass populations in Chesapeake Bay. *Copeia* **1990**: 355–366.
- Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000). Considering evolutionary processes in conservation biology. *Trends Ecol Evol* **15**: 290–295.
- Crawford AJ (2003). Huge populations and old species of Costa Rican and Panamanian dirt frogs inferred from mitochondrial and nuclear gene sequences. *Mol Ecol* **12**: 2525–2540.
- Dorazio RM, Hattala KA, McCollough CB, Skjveland JE (1994). Tag recovery estimates of migration of striped bass from spawning areas of the Chesapeake Bay. *Trans Am Fish Soc* **123**: 950–963.
- Efron B, Tibshirani RJ (1993). *An Introduction to the Bootstrap*. Chapman & Hall: New York.
- Ellegren H (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* **16**: 551–558.
- Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* **11**: 1591–1604.
- Fraser DJ, Bernatchez L (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol Ecol* **10**: 2741–2752.
- Girman DJ, Vila C, Geffen E, Creel S, Mills MGL, McNutt JW et al (2001). Patterns of population subdivision, gene flow and genetic variability in the African wild dog (*Lycaon pictus*). *Mol Ecol* **10**: 1703–1723.
- Gold JR, Turner TF (2002). Population structure of the red drum (*Sciaenops ocellatus*) in the northern Gulf of Mexico, as inferred from variation in nuclear-encoded microsatellites. *Mar Biol* **140**: 249–265.
- Goodman SJ (1997). Rst Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Mol Ecol* **6**: 881–885.
- Hamilton MB (1997). Genetic fingerprint inferred population subdivision and spatial genetic tests for isolation by distance and adaptation in the coastal plant *Limonium carolinianum*. *Evolution* **51**: 1457–1468.
- Hamilton MB, Miller JR (2002). Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. *Genetics* **162**: 1897–1909.
- Han K, Ely B (2002). Use of AFLP Analyses to assess Genetic Variation in *Morone* and *Thunnus* species. *Mar Biotechnol* **4**: 141–145.
- Han K, Li L, Leclerc GM, Hays AM, Ely B (2000). Isolation and characterization of microsatellite loci for striped bass (*Morone saxatilis*). *Mar Biotechnol* **2**: 405–408.
- Johnson JA, Toepfer JE, Dunn PO (2003). Contrasting mitochondrial and microsatellite population structure in fragmented populations of greater prairie-chickens. *Mol Ecol* **12**: 3335–3347.
- Laughlin TF, Turner BJ (1996). Hypervariable DNA markers reveal high genetic variability within striped bass populations of the lower Chesapeake Bay. *Trans Am Fish Soc* **125**: 49–55.
- Lewis PO, Zaykin D (2001). *Genetic data analysis: computer program for the analysis of allelic data*. Version 1.0 (d16c). Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>.
- Maryland Department of Natural Resources (MD DNR) (2002). *Investigation of striped bass in Chesapeake Bay*. Annual report, USFWS Federal Aid Project F-42-R-12.
- Monteleone DM, Houde ED (1990). Influence of maternal size on survival and growth of striped bass *Morone saxatilis* Walbaum eggs and larvae. *J Exp Mar Biol Ecol* **140**: 1–11.
- Morgan II RP, Koo TSY, Krantz GE (1973). Electrophoretic determination of populations of the striped bass, *Morone saxatilis*, in the upper Chesapeake Bay. *Trans Am Fish Soc* **102**: 21–32.
- Moritz C (1994). Defining 'evolutionarily significant units' for conservation. *Trends Ecol Evol* **9**: 373–375.
- Moritz C, Lavery S, Slade R (1995). Using allele frequency and phylogeny to define units for conservation and management. In: Nielsen JL, Powers GA (eds) *Evolution and the Aquatic Ecosystem: Defining Unique Units in Population Conservation*. American Fisheries Society: Bethesda, MD. Symposium 17, pp 249–262.
- Nyakaana S, Arctander P (1999). Population genetic structure of the African elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male-biased gene flow. *Mol Ecol* **8**: 1105–1115.
- Oddou-Muratorio S, Petit RJ, Le Guerroue B, Guesnet D, Demesure B (2001). Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution* **55**: 1123–1135.
- Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998). What molecules can tell us about populations: choosing and using a molecular markers. *Ecology* **79**: 361–382.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumencism. *J Hered* **86**: 248–249.
- Rice WR (1989). Analyzing tables of statistical tests. *Evolution* **43**: 223–225.
- Richards RA, Rago PJ (1999). A case history of effective fishery management: Chesapeake Bay striped bass. *N Am J Fish Manage* **19**: 356–375.
- Rosenberg NA, Burke T, Elo K, Feldman M, Freidlin PJ, Groenen MAM et al (2001). Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159**: 699–713.
- Roy NK, Maceda L, Wirgin I (2000). Isolation of microsatellites in striped bass *Morone saxatilis* (Teleostei) and their preliminary use in population identification. *Mol Ecol* **9**: 827–829.

- Rulifson RA, Dadswell MJ (1995). Life history and population characteristics of striped bass in Atlantic Canada. *Trans Am Fish Soc* **124**: 477–507.
- Ryder OA (1986). Species conservation and systematics: the dilemma of subspecies. *Trends Ecol Evol* **1**: 9–10.
- Schneider S, Roessli D, Excoffier L (2000). *Arlequin: a software for population genetics data analysis*. Ver 2.001. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.
- Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- Sidell BD, Otto RG, Powers DA (1980). Apparent genetic homogeneity of spawning striped bass in the upper Chesapeake Bay. *Trans Am Fish Soc* **109**: 99–107.
- Utter F, Ryman N (1993). Genetic markers and mixed stock fisheries. *Fisheries* **18**: 11–21.
- Waldman JR, Dunning DJ, Ross QE, Mattson MT (1990). Range dynamics of Hudson River striped bass along the Atlantic coast. *Trans Am Fish Soc* **119**: 910–919.
- Waldman JR, Richards RA, Schill WB, Wirgin I, Fabrizio MC (1997). An empirical comparison of stock identification techniques applied to striped bass. *Trans Am Fish Soc* **126**: 369–385.
- Waples RS (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *J Hered* **89**: 438–450.
- Wayne RK, Morin PK (2004). Conservation genetics in the new molecular age. *Front Ecol Environ* **2**: 89–97.
- Weir BS (1996). *Genetic Data Analysis II*. Sinauer Associates: Sunderland, MA.
- Wirgin I, Silverstein P, Grossfield J (1990). Restriction endonuclease analysis of striped bass mitochondrial DNA: the Atlantic coast migratory stock. *Am Fish Soc Symp* **7**: 475–491.
- Wirgin II, Grunwald C, Garte SJ, Mesing C (1991). Use of DNA fingerprinting in the identification and management of a striped bass population in the southeastern United States. *Trans Am Fish Soc* **120**: 273–281.
- Wirgin II, Waldman JR, Maceda L, Stabile J, Vecchio VJ (1997). Mixed-stock analysis of Atlantic coast striped bass (*Morone saxatilis*) using nuclear DNA and mitochondrial DNA markers. *Canad J Fish Aquatic Sci* **54**: 2814–2826.
- Zastrow CE, Houde ED, Saunders EH (1989). Quality of striped bass (*Morone saxatilis*) eggs in relation to river source and female weight. In: Blaxter JHS, Gamble JC, Westernhagen H (eds) *The Early Life History of Fish*. The Third ICES Symposium, Bergen, 3–5 Rapp. PV Reun. Ciem. International Council for the Exploration of the Seas (ICES): Copenhagen, Denmark. Vol 191, pp 34–42.