# Population Genetics and Dynamics of Spotted Seatrout in the Estuarine Waters of Texas

# Joel D Anderson\*, William J Karel

Texas Parks and Wildlife, Perry R Bass Marine Fisheries Research Station, 3864 FM 3280, Palacios, TX 77465, USA

\*Correspondence to: Joel D Anderson, Joel.Anderson@tpwd.state.tx.us

Published online: March 6, 2010

#### Abstract

A geospatial database of spotted seatrout (*Cynoscion nebulosus*) abundance along the Texas shoreline was generated, using 33 years of intensive gill net sampling data by the Texas Parks and Wildlife Department. The resulting landscape demographic data was used to identify areas of low individual abundance, which were evaluated as putative subpopulation boundaries. Quantitative population genetic analysis was conducted using a recently published genetic data set of mitochondrial DNA haplotypes, and a new data set of nuclear microsatellites. A significant mtDNA AMOVA result was obtained when samples were pooled into northern, mid-coast and southern spatial partitions based upon geospatial data ( $F_{ct} = 0.030$ , P = 0.003). This three-partition model captured more genetic divergence at the group level than any other model examined. Similarly, the three-partition model resulted in a small but significant group association in the microsatellite AMOVA ( $F_{ct} = 0.003$ , P = 0.009). Multivariate cluster analyses of both data sets indicated that at least three distinctive subpopulations of spotted seatrout exist within Texas waters. The genetic data are consistent with previous studies that indicate the presence of distinctive subpopulations of spotted seatrout exist within Texas coastal waters, and with geospatial abundance data indicating areas of low abundance between adjacent subpopulations along the northern and southern Texas coastline.

Keywords: Spotted seatrout; Cynoscion nebulosus; mtDNA; microsatellites; Gulf of Mexico; geospatial.

#### Introduction

The spotted seatrout (*Cynoscion nebulosus* Cuvier) is the most popular recreational marine finfish in Texas [1] and is common throughout the Gulf of Mexico (hereafter Gulf). As such, the Texas Parks and Wildlife Department manage this species intensively, and management includes annual large-scale stock enhancement in four Texas bays (Lower Laguna Madre, Upper Laguna Madre, Galveston Bay and Sabine Lake). Because of the direct and intensive management of this species, it is imperative that management goals account for the possibility of discreet subpopulations (genetic structure) within Texas waters. Previous genetic examinations of population structure of spotted seatrout in the northern Gulf, and along the Texas coast in particular, have resulted in two general models of spotted seatrout genetic structure within the Gulf.

Weinstein and Yerger [2] used allozyme data to characterize genetic diversity in spotted seatrout from the Gulf coast of Florida, and concluded that spotted seatrout were organized into discreet subpopulations. Gold et al. [3] used the distribution of mitochondrial DNA (mtDNA) RFLP haplotypes to conclude that samples taken of spotted seatrout throughout the northern Gulf and Atlantic (including samples from Texas) generally represented distinct subpopulations, and suggested that these subpopulations may be organized by natal estuaries, as expected from short dispersal distances in tagging studies and the affinity of spotted seatrout for inshore areas. In contrast with the finding of discreet subpopulations, Ramsey and Wakeman [4] and King and Pate [5] used allozymes and found low measures of population subdivision ( $F_{st}$ ) and high gene flow ( $N_{e}m$ ) in Gulf samples, indicating that individuals in the western/northern Gulf

comprise a single large population. Both studies found isolation-by-distance (IBD) among locales, suggesting that genetic differentiation occurred on a broad (but continuous) geographic scale. Recently, the authors of this study examined mtDNA control region haplotype data and reached a similar conclusion of IBD; evidence for discreet subpopulations was weak, and driven primarily by divergence at the extreme ends of the sampling distribution [6]. In any event, the effective genetic differences between distinctive subpopulations, versus a single population operating under IBD, are likely to be subtle because the genetic drift that is expected among distinctive subpopulations can potentially be offset by the large census sizes that generally characterize marine finfish populations. Thus, Gold et al. [7] suggested that although subpopulations centered in natal estuaries may have distinctive short-term demographic characteristics, recurrent gene flow among adjacent estuaries results in a genetic pattern which resembles IBD, rather than distinct subpopulations.

In all previous studies, an empirical method for formulating hypotheses regarding the geographic size and demography of putative spotted seatrout subpopulations has not yet been employed. These studies have generally invoked the assumption that subpopulations will most likely be geographically centered within inshore bays and estuaries. However, Anderson and Karel [6] pointed out that there was little physical structure available that could act as barriers to limit migration among most inshore areas in Texas. Thus, the spatial component of designated estuaries along the coast of Texas amount to little more than political designations, rather than disparate entities with completely independent geography. In this study, a novel approach to defining putative subpopulation boundaries was used to develop hypotheses regarding the genetic structure of spotted seatrout in Texas. A geospatial data set representing the abundance of spotted seatrout in Texas' inshore coastal areas was extracted from a fishery-independent database of 33 years of Texas Parks and Wildlife Department gill net sampling. These data were used to construct a map of spotted seatrout abundance along the entire Texas coast, which resulted in a more systematic basis upon which putative subpopulation boundaries could be formulated. That is, subpopulation boundaries might be expected to occur in areas where spotted seatrout abundance was relatively low. These putative boundaries were then evaluated quantitatively by genetic data analysis. Two genetic data sets were examined: 1) the mtDNA sequence data of Anderson and Karel [6] was reanalyzed, and 2) a new data set of microsatellite markers was added. The results of the current study were evaluated in the context of all other previous studies [2-8], and were used to add resolution to the question of population structure of spotted seatrout on the coast of Texas.

#### Methods

# Gill net sampling and geospatial data

Spatial data was generated by collating standardized gill net catch data from 1.85 km-square shoreline grids covering the extent of the Texas coast. Gill nets have been deployed biannually coast-wide for 33 years, during fishery-independent departmental monitoring by the Texas Parks and Wildlife Coastal Fisheries Division (hereafter TPWD). Each major estuarine system has been sampled weekly, for a 10 week period during the spring and a 10 week period during the fall. Grids for sampling are chosen at random with 3-5 grids sampled per week, and no grid is sampled twice in the same month. Gill nets, with mesh sizes ranging from 76 mm to 152 mm, were 184 m long and partitioned into four 46 m sections. This data set comprised 22,824 data points covering the extent of the Texas shoreline for years 1975-2008, including the bay side of barrier islands and peninsulas.

For each grid, catch-per-unit-effort (CPUE) was assessed by averaging multiple independent records of CPUE over time. Averaged CPUE was then rounded to the nearest whole integer, and each grid was categorized as having a CPUE of 0, 1, 2 or 3+ over the entire sampling period. Category scores were then superimposed as abundance-graded circles over grids on the Texas shoreline using ArcMap v9.3 (ESRI, Redlands, CA, USA). The abundance data was examined (qualitatively) for areas where multiple adjacent grids had low average abundance (CPUE = 0). These areas were hypothesized to be putative subpopulation boundaries, which were evaluated quantitatively using genetic methods (see below).

## Mitochondrial DNA sequence data

The data set of Anderson and Karel [6] was reanalyzed here to examine patterns of mtDNA control region haplotype distribution which correlated with geospatial data. Briefly, this data set comprised young-of-the-year (YOY) spotted seatrout from eight of Texas' nine major bay systems over a three year period (hereafter, SL = Sabine Lake; EM = East Matagorda Bay; WM = West Matagorda Bay; SB = San Antonio Bay; AB = Aransas Bay; CC = Corpus Christi Bay; UL = Upper Laguna Madre; LL = Lower Laguna Madre, Fig. 1). Within each bay, care was taken to avoid sampling possible family groups, by taking individuals from multiple sampling locales on multiple dates. Samples consisted of either >200 mg fin clips, or whole fish in the case of small individuals. Samples were preserved and processed according to Anderson and Karel [6].

Genetic differentiation among groups and among samples (localities) within groups was tested with an analysis of molecular variance (AMOVA, [9]). Individual samples were pooled across years and variance in mtDNA haplotype association was partitioned into between-individual, among-samples and among-group effects using ARLEQUIN v3.0 [10], using simple pair wise differences. Four different models of population structure were tested, as follows: 1) all sampling locales grouped as single subpopulations (SL) + (EM) + (WM) + (SB) + (AB) + (CC) + (UL) + (LL), 2) three subpopulations consisting of (SL) + (EM, WM, SB, AB, CC) + (UL, LL), 3) three subpopulations consisting of (SL) + (EM, WM, SB, AB, CC, UL) + (LL), and 4) four subpopulations consisting of (SL) + (EM, WM, SB, AB, CC) + (UL) + (LL). These models were chosen based on the low abundance of spotted seatrout in areas between LL and UL, between UL and CC, and between EM and SL, as indicated from examination of the geospatial data set (see results). A Mantel matrix correlation analysis was also used to examine isolation-by-distance effects. This Mantel procedure was identical to that of Anderson and Karel [6] and is only presented here for comparison purposes to a similar Mantel matrix test on the microsatellite data.

Finally, a hierarchical cluster analysis was performed on pairwise DNA distance data between samples, in order to determine the number of groups (subpopulations) present in the data set, and to simultaneously determine group membership. Nei's corrected genetic distance [11] was generated between each pair wise set of samples. The dissimilarity matrix of Nei's values was then subjected to clustering using Ward's minimum variance linkage method [12]. This method was chosen because it minimizes within-group dispersion and is effective in eliciting among-cluster relationships [13], which is preferred in situations where clusters are expected to be defined by short distance. The number of clusters to retain was determined qualitatively by using an elbow-plot of genetic distance values; that is, the expected group number was equal to the number of clusters (*k*) present as the genetic distance accounted for by each successive addition of clusters leveled off. Thus, within-group dispersion minimization was not improved with additional levels of clustering, once the optimum number of clusters had been obtained. Clustering was conducted using JMP 7.0.1 software (SAS Institute, Cary, NC, USA). Statistical support for major nodes on the cluster dendrogram was obtained by bootstrap analysis using PHYLIP 3.63 software [14], with 100 bootstrap replications over geographic samples.

# Microsatellite marker data

For the microsatellite data set, spotted seatrout YOY were collected over a three-year period from the same eight geographic localities as those from the mtDNA data set. In some, but not all cases, samples from the mtDNA and microsatellite data sets overlapped. Each data set was collected and analyzed at different times; although congruence in samples used was attempted, in some cases samples were no longer available or were degraded. For this reason, each data set will be treated independently. Microsatellite samples were taken as either whole fish or as fin clips of larger

fish and submerged immediately in vials containing 95% ethanol (EtOH). The vials were placed on ice or stored at 4°C for 24 hours prior to shipment to the Perry R. Bass Marine Fisheries Research Station (PRBMFRS).



FIGURE 1: Map of the coastline of Texas, including locations where samples of Cynoscion nebulosus were taken for genetic analysis.

Once at PRBMFRS, samples were stored in ethanol at room temperature for an indeterminate period. Approximately 200 mg of tissue were used for individual DNA isolations. All DNA isolations were carried out using PUREGENE<sup>®</sup> DNA isolation kits (Gentra Systems, Inc., Minneapolis, MN), with a final rehydration volume of 75  $\mu$ l.

Six microsatellite loci were used to examine population structure in spotted seatrout (Table 1). These loci were developed from the closely related Sciaenid *Sciaenops ocellatus*, and initial characterization of these loci can be found elsewhere [15, 16]. These loci were amplified via the Polymerase Chain Reaction (PCR) under a modified touchdown protocol, using Ready-To-Go<sup>TM</sup> PCR beads (GE Healthcare, Piscataway, New Jersey) on a Techne Genius thermocycler (Techne Inc., Princeton, New Jersey). Reactions consisted of 1 µl template DNA (50 ng/µl), 1 Ready-To-Go<sup>TM</sup> bead, and 24 µl of forward and reverse primer cocktail (0.4 µM standard primer concentration of each primer), for a total of 25 µl. Each reverse primer oligonucleotide was purchased labeled with a WellRED dye (Sigma-Proligo, Boulder, Colorado) in preparation for visualizing on a Beckman-Coulter CEQ8000<sup>TM</sup> automated sequencer (Beckman Coulter Inc., Fullerton, California). The touchdown PCR protocol utilized for all reactions consisted of the following: an initial single denature period of 2 min at 95°C, 10 cycles of initial amplification (95°C for 30 s, 55°C for 30 s, and 72°C for 1 min adding 3 s of extension per cycle), and a single final extension period of 7 min at 72°C. PCR products were combined with a 400bp size standard and separated on a Beckman-Coulter CEQ8000<sup>®</sup>. Fragment analysis was performed using software provided by Beckman-Coulter and all genotypes were examined visually for accuracy.

The program FSTAT v2.9.3.2 [17] was used to measure microsatellite diversity statistics including allelic richness ( $R_{s_i}$ , [18]), gene diversity ( $H_e$ ), the coefficient of inbreeding within samples ( $F_{is}$ ), as well as the fixation index among all samples ( $\vartheta$ ). Additionally, FSTAT was used to test for significant genetic linkage between pairs of microsatellites using data randomization [17] with an arbitrary cutoff of  $\alpha = 0.05$ . The statistics  $F_{is}$  and  $\vartheta$  were determined initially using the methodology of Weir and Cockerham [19]. The standard error of the fixation index ( $\vartheta$ ) was determined by jackknifing over samples, and was used to determine whether there was statistical support for heterogeneity among samples at each locus. The coefficient of inbreeding ( $F_{is}$ ) was used to determine whether there was any indication of non-random allelic associations within samples. Randomization tests were performed with 1000 iterations in order to determine whether  $F_{is}$  differed significantly from 0.0, which could indicate deviation from Hardy-Weinberg genotypic expectations caused either by unsampled null alleles, by differential amplification of alleles or otherwise systemic errors in allele size calling for a give locus. The nominal level for statistical significance in randomizations was adjusted for multiple tests performed simultaneously ( $\alpha = 0.001$ ).

As with the mtDNA data set, microsatellites were used in a distance-based AMOVA [9]. Initial examination of this data set revealed homogeneity among cohorts sampled in different years. Cohorts from all three years were thus pooled into single samples for spatial analyses. Variance in nuclear genotype association was partitioned into between-individual, among-sample and among-group effects using ARLEQUIN v3.0. The distance metric chosen for microsatellite AMOVA was the average number of pair wise allelic differences between individual genotypes. The same four population models used in mtDNA AMOVA were tested with the microsatellite AMOVA. Additionally, a Mantel matrix correlation analysis was used to examine isolation-by-distance of microsatellites and compared to the mtDNA Mantel test. Genetic divergence between pair wise samples was estimated ( $F_{st}$ ) and subsequently compared to geographic distance in a matrix correspondence test as implemented in ARLEQUIN v3.0. The significance of the regression coefficient of matrix correspondence (r) was assessed by a 1000 iterations of a randomization procedure, as implemented in ARLEQUIN v3.0. Finally, a cluster analysis was performed in order to simultaneously test for number of groups (subpopulations) and group membership. For the microsatellite cluster analysis, the Cavalli-Sforza chord measure [20] was generated using PHYLIP 3.63, and the dissimilarity matrix of chord distances was subjected to Ward's minimum-variance linkage clustering. As with the mtDNA data set, clustering was carried out using the JMP 7.0.1 software, and statistical support for nodes was obtained by bootstrap resampling in PHYLIP 3.63.

TABLE 1: Microsatellite loci statistics, including gene diversity (H<sub>2</sub>), allelic richness (R<sub>2</sub>) and the coefficient of inbreeding (F<sub>2</sub>) in each sample of spotted seatrout, and overall. The "alleles" column indicates the number of alleles sampled, overall, for each locus. The value & was jack-knifted over samples to determine whether there was significant ( $\pm$ ) partitioning of genetic variance at each locus.

			Bundhare	area				200				
Locus	Atteles	Diversity	SL (n=41)	(n = 67)	(n = 78)	(n = 119)	(n = 119)	(n = 79)	(n = 41)	(n = 12)	Overall	8 (among pops)
soc12	10	He	150	0.47	0.55	0.49	0.52	0.52	0.51	0.38	0.49	-0.002 (± 0.003)
		Ŗ	4.63	3.08	3.88	325	3.46	3.79	3.42	5.00	3.60	
		$\mathbf{F}_{\omega}$	0.09	0.02	0.02	-0.04	90.06	-0.15	0.24*	11.0-	1	
80050	62	$H_{\epsilon}$	0.59	0.57	0.55	0.59	0.59	0.57	0.61	0.55	0.58	0.005 (± 0.002)*
		R,	328	328	2.90	3.46	3.48	3.10	3.48	3.00	3.30	
		Fa	0.05	0.00	0.15	0.08	-0.01	0.14	-0.20	-0.06	1	
soc243	Ľ	H,	0.46	0.51	0.43	0.52	050	0.48	0.49	0.46	0.48	-0.002 (± 0.002)
		R,	2.88	2.81	2.31	3.10	3.04	2.28	3.00	2.00	2.77	
		Fa.	-0.06	-0.08	0.14	-0.12	0.02	-0.08	-0.06	-0.08	C	
soc412	0	$H_{\epsilon}$	0.42	0.51	0.36	0.45	0.42	0.35	0.23	0.16	0.36	0.004 (± 0.005)
		R.	4.14	4.96	3.87	4.55	4.05	3.92	335	2.00	4.16	
		F.	0.06	-0.03	-0.04	+110	10.0-	-0.03	0.15	-0.05	1	
soc415	38	H	500	0.92	0.92	160	0.92	16.0	760	0.88	16.0	-0.001 (± 0.001)
		R,	13.49	12.58	12.40	11.70	12.32	12.50	12.42	10.00	12.37	
		Fa.	0.01	10'0	0.05	-0.01	0.07+	-0.02	10.0-	0.14	1	
soc432	19	$H_{\epsilon}$	0.73	0.74	0.73	0.73	0.72	0.67	0.76	0.61	0.71	0.000 (± 0.002)
		K.	622	727	6.36	6.87	6.76	6.86	738	4.00	6.81	
		10	-0.14	+51.0	£0.0	-0.08	0.03	0.04	-0.10	0.04	1	

http://astonjournals.com/faj

No values of Fa were significant after Bonferroni correction.

denotes significance of F<sub>a</sub> prior to correction for multiple tests conducted simultaneously.

## Results

## Spatial abundance of spotted seatrout in Texas

Spotted seatrout occurred ubiquitously along the Texas shoreline during the duration of TPWD sampling. Of 1,806 grids sampled, 1,785 contained spotted seatrout on at least one occasion, for an overall occurrence ratio of 0.99 (Table 2). Average CPUE among all estuaries was 0.58 ( $\pm$  0.12) and ranged from 0.21 (SL) to 0.94 (LL). It is important to note that heterogeneity between spring and fall gill net spotted seatrout CPUE has been anecdotally reported, but was not explicitly examined here. The coast-wide CPUE of 0.58 represents the average among all grids, in both seasons.

Qualitative observation of localized abundance indicated five areas of relatively low or ambiguous spotted seatrout abundance. The first area (area #1, Fig. 2a) was in the upper coast between SL and the unsampled Galveston bay (GB). In this area, there is not a substantial natural inshore waterway or bay area, and the Gulf Intracoastal Waterway (ICWW) represents the only inshore route of migration between SL and GB. The routine monitoring of TPWD does not set gill nets in the ICWW, so there is no estimation of abundance in this area. However, the absence of natural inshore waterways represents a physical bottleneck, and for the purpose of this project, it will be assumed such areas represent a possible dispersal or migration barrier. A second area of low/ambiguous abundance occurs between GB and EM (area #2, Fig. 2a). This is another area where inshore waterways are limited, and the ICWW represents the only substantial dispersal route for inshore species. Because GB was not sampled in this study, boundary areas #1 and #2 will be considered interchangeable for the purpose of this study, and limitations of genetic dispersal in this area will be tested by assuming that the SL sample represents a distinct subpopulation.

Between EM and CC (i.e. "mid-coast"), there is little evidence of low abundance in any single area (Fig. 2b). The midcoast area is characterized by large natural waterways that act as corridors between adjacent bays, in addition to the ICWW. Spotted seatrout abundance in this area is homogeneous and thus no putative subpopulation boundaries were tested in this area. In contrast, from CC south to LL there are three areas of low/ambiguous abundance. The first (area #3, Fig. 2b, 2c) is located at the northern end of UL. Despite a large natural inshore waterway between UL and CC, this area has the lowest overall abundance of any inshore area measured, and contains several adjacent grids where CPUE rounded down to 0.0. The northern UL is characterized by hyper salinity most of the year (as is the entire Laguna Madre) and extremely shallow water (~ 1-2 m, on average). Limitations of genetic dispersal in this area will be tested by assuming that Laguna Madre samples (UL + LL) together form a distinct subpopulation. A fourth low/ambiguous abundance area is between UL and LL (area #4, Fig. 2c). This is another area where there is no natural waterway, and the ICWW represents the only means of dispersal. Limitations of genetic dispersal in this area will be tested by assuming that the LL sample is its own distinct subpopulation. Finally, multiple localized areas of low abundance are present in the LL (area #5, Fig. 2c). Independent tests of divergence among multiple LL samples was not possible because sample sizes from this bay were small (n = 27 mtDNA, n = 12 microsatellite DNA) and were distributed around the entire system.

#### Mitochondrial DNA haplotype data

The data set used in this study was identical to the data from [6], however a brief overview of the data will be given here. Sample sizes ranged from 27 individuals (LL, WM) to 44 individuals (SL, Table 3) sequenced for a portion of the mtDNA control region. The overall mtDNA sequence set included 280 examined individuals, sequenced over 373 nucleotides, and resulted in recovery of 60 novel haplotypes. There was no evidence for heterogeneity in base frequencies among haplotypes, and the rate of transitions/transversions across a range of pair wise genetic distances indicated no evidence for mutation saturation. Among sampling sites, haplotype diversity ( $H_d$ ) ranged from 0.754 (SL) to 0.940 (WM) (Table 3). Nucleotide diversity ranged from 0.005 (SL) to 0.007 (EM, WM, SB, CC, LL). There was no indication of heterogeneity among annual samples taken at the same location, thus cohorts were combined for spatial analyses. The AMOVA procedure was used to test group (subpopulation) structure of spotted seatrout samples, using four different models derived from qualitative examination of spotted seatrout abundance (Table 4a). The model which assumed that each sampled bay harbored its own distinctive spotted seatrout subpopulation resulted in a small and statistically non-significant value of among-group structure ( $\Phi_{st} = 0.011$ , p = 0.095). All other models, in which the mid-coast bays (EM through CC) were grouped together, resulted in significant among-group structure. Of these, the model (SL) + (EM, WM, SB, AB, CC) + (UL, LL) captured the highest among-group divergence ( $\Phi_{st} = 0.030$ ) and achieved the highest statistical support (p = 0.003). This model assumed that the SL sample and both Laguna Madre samples (UL+LL, combined) came from independent subpopulations (all mid-coast samples came from a single population). Additionally, there was significant evidence of IBD, coast wide. The Mantel test indicated a strong and significant correlation between geographic distance (km) and genetic divergence ( $\Phi_{st}$ ) of mtDNA haplotypes among samples (r = 0.717, p = 0.010).

TABLE 2: Abundance data for spotted seatrout samples from Texas. The number of grids sampled represents the raw number of actual data grids that were queried in the 33 years of fishery-independent sampling (each grid represents 1 km<sup>2</sup> of inshore shoreline waters). The positive (+) grids and occurrence ratio indicate the number and ratio of grids sampled that contained spotted seatrout on at least one occasion. The averaged CPUE is the per-sample catch-per-unit-effort of spotted seatrout averaged over all grids, for every observation.

Area	Code	Grids	sampled	(+) grids	Occurrence	Averaged CPUE (± SE)
		( <i>n</i> )			ratio	
Sabine Lake	SL	91		85	0.93	0.21 (± 0.03)
Galveston Bay	GB	290		288	0.99	0.54 (± 0.05)
East Matagorda	EM	66		61	0.92	0.57 (± 0.08)
West Matagorda	WM	422		420	1.00	0.62 (± 0.03)
San Antonio Bay	SB	327		326	1.00	0.64 (± 0.03)
Aransas Bay	AB	201		199	0.99	0.55 (± 0.05)
Corpus Christi Bay	CC	106		106	1.00	0.64 (± 0.12)
Upper Laguna Madre	UL	183		183	1.00	0.49 (± 0.05)
Lower Laguna Madre	LL	120		117	0.98	0.94 (± 0.12)
Overall		1806		1785	0.99	0.58 (± 0.12)



FIGURE 2: Relative abundance of *Cynoscion nebulosus* in the upper coast (a), middle coast (b) and lower coast (c) of Texas. Abundance is measured as catch-per-unit-effort (CPUE), rounded to the nearest whole integer, and is graded by circle size. Areas with dots rather than circles (Sabine Lake, Upper Laguna Madre) had grids with CPUE that rounded down to 0.0. The five boxed numbers (1-5) indicate areas of low or ambiguous abundance (as discussed in the text), which were assumed to be potential subpopulation boundary areas.

Area	N (bay)	Haplotypes	H <sub>d</sub>	π
SL	44	12	0.75	0.005
EM	28	15	0.92	0.007
WM	27	16	0.94	0.007
SB	40	18	0.90	0.007
AB	42	17	0.89	0.006
CC	42	16	0.89	0.007
UL	30	12	0.87	0.006
LL	27	12	0.89	0.007

TABLE 3: Sample sizes and mtDNA diversity indices for spotted seatrout samples. Sample data includes the overall sample size (N), the number of haplotypes recovered, haplotype diversity ( $H_d$ ) and nucleotide diversity ( $\pi$ ).

Multivariate cluster analysis also indicated three groups of samples. Qualitative assessment of the elbow criteria indicated that the level of clustering containing three main clusters accounted for 64% of the genetic distance assayed (Fig. 3). The first sample group to separate from all others included only the SL sample. In the second level of clustering, the UL and LL samples separated from all others, leaving the final mid-coast grouping of EM, WM, SB, AB and CC. These groupings were consistent with those that achieved the highest and most significant genetic divergence in AMOVA analyses. Bootstrap support for the mid-coast and Laguna Madre nodes were 100%.

#### Microsatellite DNA data

For the microsatellite dataset the number of alleles per locus ranged from seven (soc243) to 38 alleles (soc415, Table 1). Overall allelic richness (relative variability corrected for sample size) ranged from  $R_s = 2.77$  (soc243) to  $R_s = 12.37$  (soc415), while overall gene diversity ranged from  $H_e = 0.36$  (soc412) to  $H_e = 0.91$  (soc415). In four cases, there was evidence for single locus deviation from Hardy Weinberg expectations, in single populations, as demonstrated by significantly high (> 0) values of  $F_{is}$ . However, in each case these values were not significant following correction for multiple tests run simultaneously. Nor was there significant evidence for linkage between any two pairs of loci. Finally, there was little evidence for heterogeneity in allele distributions among samples at individual loci. Only a single locus (soc50) had a fixation index among populations which was significantly different from 0.0 ( $\vartheta = 0.005 \pm 0.002$ ).



TABLE 4: AMOVA results for among-group divergence ( $\phi_{st}$  = mtDNA,  $F_{st}$  = microsatellite DNA) for four models of subpopulation structure of spotted seatrout on the Texas coast. In each model, numbers represent the numerical index and membership of each predefined subpopulation.

_	Model subpopulations	$\Phi_{st}$	Р
	1. SL	0.021	0.020
	2. EM, WM, SB, AB, CC, UL		
	3. LL		
	1. SL	0.030	0.003
	2. EM, WM, SB, AB, CC		
	3. UL, LL		
	1. SL	0.023	0.017
	2. EM, WM, SB, AB, CC		
	3. UL		
	4. LL		
	All bays independent (8 subpops)	0.011	0.095
	B. Nuclear microsatellites		
_			
	Model subpopulations	<i>F</i> <sub>st</sub>	Р
	1. SL	0.004	0.033
	2. EM, WM, SB, AB, CC, UL		
	3. LL		
	1. SL	0.003	0.015
	2. EM, WM, SB, AB, CC		
	3. UL, LL		
	1. SL	0.003	0.025
	2. EM, WM, SB, AB, CC		
	3. UL		
	4. LL		
_	All bays independent (8 subpops)	0.001	1.000



FIGURE 3: Cluster analysis of mtDNA haplotypes, based on Nei's corrected genetic distance [11]. The dendrogram represents the outcome of multivariate cluster analysis using Ward's (1963) minimum-variance linkage clustering. Numbers at nodes are statistical support for each grouping, obtained by bootstrap resampling of haplotype frequencies (only nodes  $\geq$  50% are shown). The graph below the dendrogram is the elbow-graph of genetic distance by number of clusters, and lines from the graph indicate group membership at each cluster level. The red line represents the number of clusters retained based on the elbow criterion (qualitative), which minimized within-group dispersion.

As in mtDNA analyses, samples were grouped using the four different models derived from qualitative examination of spotted seatrout abundance (Table 4b). The model which assumed that each sampled bay harbored its own distinctive spotted seatrout subpopulation resulted in a small and statistically non-significant value of among-group structure ( $F_{st} = 0.001$ , p = 1.000), which was further supported by a statistically significant fixation index at only a single locus (soc50, Table 1). In addition, similar to mtDNA data, all other models tested resulted in significant among-group structure. Of these, the model (SL) + (EM, WM, SB, AB, CC) + (UL, LL) captured the 2<sup>nd</sup> highest among-group divergence and achieved the highest statistical support ( $F_{st} = 0.003$ , p = 0.015). The model which considered UL to be part of the mid-coast subpopulation (SL) + (EM, WM, SB, AB, CC, UL) + (LL) achieved the highest level of among-group divergence, but had relatively weak statistical support ( $F_{st} = 0.004$ , p = 0.033). One possible explanation for the weak statistical support of this model is that the LL sample was the smallest of all samples taken (n = 12); this result should therefore be treated with caution.



FIGURE 4: Cluster analysis of microsatellites, based on chord distance. The dendrogram represents the outcome of multivariate cluster analysis using Ward's (1963) minimum-variance linkage clustering. Numbers at nodes are statistical support for each grouping, obtained by bootstrap resampling of haplotype frequencies (only nodes  $\geq$  50% are shown). The graph below the dendrogram is the elbow-graph of genetic distance by number of clusters, and lines from the graph indicate group membership at each cluster level. The red line represents the number of clusters retained based on the elbow criterion (qualitative), which minimized within-group dispersion.

Multivariate cluster analysis indicated the presence of three groups using microsatellite data. These three groups accounted for 75% of the genetic distance measured (Fig. 4). However, in contrast to the mtDNA data, microsatellite data indicated that the UL sample separated from all others at the first level of clustering. At the second level of clustering, the SL sample separated from all others, leaving a final grouping of EM, WM, SB, AB, CC and LL, with 95% bootstrap support. Heterogeneity in genetic dispersion within groups leveled off after the second cluster level.

Similar to the mtDNA data, there was significant evidence of IBD from the microsatellite data. The Mantel test of microsatellite genotype data indicated a strong and significant correlation between geographic distance (km) and genetic divergence ( $F_{st}$ ) among samples (r = 0.617, p = 0.034).

#### Discussion

The main intent of this study was to address two questions: 1) are there areas of chronically low abundance of spotted seatrout on the Texas coast, and 2) do spotted seatrout samples among these areas demonstrate significant discontinuity in the distribution of genetic variability. The data from this study indicate that there are areas along the Texas shoreline that indeed have chronically low abundance of spotted seatrout (such as the northern UL). Additionally, inshore waterways between adjacent bays on the northern (SL, GB) and southern (UL, LL) Texas coast consist only of the ICWW, which is of ambiguous suitability for sustaining high levels of gene flow between distant estuaries. While the middle-coast area of Texas (EM – CC) demonstrates continuously high abundance from north to south, sampling areas in the northern (SL) and southern (UL, LL) extremes resulted in higher spatial variability of spotted seatrout abundance. The low abundance of spotted seatrout in the northern UL, as well as the lack of natural inshore waterways north and south of GB, and between UL and LL, could represent areas where bay-to-bay dispersal is relatively low. For this reason, these areas may be intuitively regarded as possible subpopulation boundary areas. However, one possible caveat to this finding is that the data that was used to characterize putative subpopulation boundaries was limited in scope to areas that could support deployment of a large gill net. Thus, the differences in geospatial abundance described by this data might also be influenced by limitations of the TPWD routine monitoring sampling strategy.

Regarding the second question of discontinuity in the distribution of genetic variability among regions, results from both mtDNA control region sequences and microsatellites suggest the potential for multiple subpopulations of spotted seatrout in Texas' coastal waters. Spatial abundance analysis results in a more systematic assessment of testable subpopulation boundaries, which in turn results in more statistically robust AMOVA results than those attained in the previous study by the authors using the same data [6]. Additionally, the cluster analyses, in contrast to AMOVA, allowed group membership to be quantitatively assigned. In particular, the benefit of multivariate clustering is that the number of groups (or subpopulations, *k*) and the sample membership in each group can be simultaneously assessed, without the need for group assignment *a priori*. Both mtDNA and microsatellite cluster analyses indicate subpopulation structure that is generally consistent with landscape abundance results. This contradicts the previous assertion that significant genetic distances at the extreme ends of a single IBD population alone were driving significant AMOVA results for Texas spotted seatrout [6].

Operating under the assumption that discreet subpopulations of spotted seatrout may exist within the territorial waters of Texas, microsatellite and mtDNA data are consistent on three points. First, the SL sample, which is the northernmost sample location, is divergent from southern samples. Both markers indicate significance of among-subpopulation AMOVA when SL is treated independently, and SL is indicated in cluster analyses as being independent from other samples. Second, there is some indication that UL and LL, the southernmost groups, also represent discreet subpopulation(s). However, the composition of subpopulation structure in the Laguna Madre samples differs between marker types. Evidence from mtDNA suggests that the UL and LL make up a single discreet subpopulation, independent

from all others. This assertion is supported both by AMOVA and cluster analysis. In contrast, while microsatellite AMOVA supports the continuity of the Laguna Madre samples (UL+LL), cluster analysis indicates that the UL sample is independent of the LL sample, which clusters with mid-coast bays. One problem with this result is the microsatellite sample size for this sample; the LL microsatellite sample was by far the smallest sample taken in this study (n = 12). The Cavalli-Sforza chord measure distance method used for clustering microsatellites is allele frequency-based [20]. Therefore, it is likely that the small sample size of LL resulted in poor estimation of allele frequencies in the LL sample, which in turn resulted in ambiguous results in the cluster analysis. The mtDNA sample taken for this site was much larger (n = 27) and indicated continuity of the Laguna Madre. Finally, agreement between the two data sets indicates continuity of mid-coast regions, from EM to CC. In the case of clustering, the mid-coast samples all clustered together in the final level of clustering, whereas no genetic distances among mid-coast samples were significant in pair wise tests.

The magnitude of genetic divergence measured among putative subpopulations of spotted seatrout in this study is much higher as measured by mtDNA than that measured by microsatellite DNA. The model-based divergence between subpopulations as measured with mtDNA haplotypes (0.021 <  $\phi$  <0.030) is an order of magnitude higher than the divergence measured with microsatellites ( $0.003 < F_{st} < 0.004$ ). The difference between nuclear and mtDNA estimates of genetic divergence has now been consistent across a number of studies. Prior microsatellite examinations of spotted seatrout were unable to obtain consistent statistical support for divergence among Gulf suppopulations [7, 8]. The magnitude of divergence among Texas populations in these studies was measured between 0.0<  $\vartheta$  < 0.015 (range across loci, [7]), and was not significant at any locus. In contrast, a prior estimate of divergence among western Gulf localities using mtDNA was  $\phi_{st} = 0.025$  (p < 0.001) [3]. The difference between nuclear and mtDNA estimates of divergence could be influenced by different genetic effective size ( $N_e$ ) or migration rate (m) among males and females. In the case of  $N_{er}$ the effective size of mtDNA is naturally one quarter of the effective size of nuclear DNA, because of the maternal origin of all mtDNA molecules (assuming 50/50 ratio of males/females) in combination with the haploid nature of mitochondrial genomes. If the ratio of males/females is skewed towards males, or if there is higher variance of reproductive success of females in comparison to males, this will result in an even smaller effective size of mtDNA (and presumably, elevated divergence). For instance, variance in reproductive success was suggested to play a role in dramatic differences between the genetic effective sizes of two closely related sciaenids in the western Gulf (C. nothus, C. arenarius [21]), and has been used to explain the disparity between effective population size and census size in the closely related *Sciaenops ocellatus* [22]. Finally, if gene flow (m) is primarily male-mediated in spotted seatrout, this will also result in a lower level of measured genetic divergence at nuclear loci. Limited female dispersal from natal areas has likely affected divergence at mtDNA loci [3]. Limited dispersal is also supported by tagging studies in Texas [23], although it is unclear whether there are differences in magnitude between male and female movements. A final possible explanation for the disparity between mtDNA and microsatellite divergence is that divergence measured by microsatellites is often influenced negatively by microsatellite polymorphism [24]. For instance, an inverse relationship between  $F_{st}$  and both allele richness and heterozygosity was previously demonstrated in the marine fish walleye pollock Theragra chalcogramma [25]. This effect could lead to uncertainty in interpreting significant population differentiation at microsatellite loci, as biological processes leading to divergence are difficult to disentangle from the evolutionary mechanics of microsatellite mutation [26].

The results of this study suggest different management implications than those of Anderson and Karel [6], which reached the conclusion that spotted seatrout in Texas were members of a single continuous population, with limited (intrinsic or extrinsic) barriers to dispersal. In an attempt to place spotted seatrout in Texas into one of three theoretical models described by Laikre et al. [27], Anderson and Karel [6] suggested that "continuous change" was more suitably supported by the data than discrete subpopulations. Indeed, strong clinal gradients in allozyme allele frequencies [5, 28] and mtDNA haplotypes [3,6] have been found in this species. However, the population criterion of Laikre et al. [27] must be viewed as existing on a continuum between no differentiation on one end, and strong allopatric isolation of distinctive groups on the other. Spotted seatrout exist somewhere in the middle of that continuum, and on local scales, the substructure of these populations is likely affected both by isolation-by-distance (IBD), as well as by some restricted

local migration among adjacent sites (Fig. 5). This is most consistent with the model hypothesized by Gold et al. [7], who suggested that spotted seatrout exist in a series of overlapping subpopulations with sufficient migration as to prevent accumulation of significant genetic divergence. The evidence from the current data further predict that there may be areas on the coast of Texas in which small-scale migration is fairly limited; it is not clear what restricts migration, but as previously mentioned there is some evidence for natal fidelity for inshore areas (primarily of females), which may be reinforced by encounters with areas of low habitat suitability (such as the northern UL).



FIGURE 5: Proposed genetic structure of spotted seatrout in Texas. Location codes are described in the text. The subpopulation labels upper-coast, mid-coast and lower-coast are based on subpopulation boundaries evaluated in this study. The hard line between CC and UL represents a boundary directly tested with genetic data. The dashed lines between EM and GB, and between GB and SL, have equal probability of driving divergence between the upper- and middle-coast; this uncertainty is the result of the Galveston Bay location (GB) being unsampled in this study. The table (inset) is direct comparisons of nuclear genetic divergence ( $\vartheta$ , microsatellites) between the three putative subpopulations.

This results in a more complicated model of population structure than one that can be adequately characterized by one of Laikre's aforementioned categories. Given this complexity, some general management implications for spotted seatrout in the western Gulf that are supported by this study, as well as previous studies, will be outlined here.

First, multiple studies agree that populations of spotted seatrout in the Laguna Madre represent a demographically independent stock. King and Pate's [5] phenogram of genetic similarity based upon allozyme data suggested that samples taken in the UL and LL region were more similar to a sample from Mexico than any samples further north. Subsequently, Gold et al. [3] indicated notable "steps" in mtDNA haplotype diversity in this area as well, and both data sets in this study support this finding. Texas Parks and Wildlife Department life history data indicate a decline in overall abundance and spawning stock biomass in LL, compared to gains in estuaries further north (unpublished data), suggesting some degree of demographic independence in the Laguna Madre. Currently, UL and LL are two of the four bays that are stocked through marine enhancement practices by TPWD. The current strategy is for stocked fingerlings to be released back into the same estuary from where the corresponding brood stock were originally extracted, although occasionally logistical considerations result in stocking into adjacent estuaries. Both of these scenarios are supported by the current data, which indicate continuity of the Laguna Madre. However, caution must be exercised, particularly as it pertains to the genetic effective size of released individuals in the LL. Gold et al. [3] reported haplotype diversity in LL that was lower than anywhere else in Texas was. This finding has not been repeated in this study or allozyme/microsatellite studies, but this result could be an indication that haplotype diversity is highly variable among years [6].

Second, multiple studies have now indicated genetic divergence between upper-coast estuaries (GB and SL), and estuaries further south. The microsatellite study of Ward et al. [8] indicated allele heterogeneity between the two northernmost bays and those occurring further south, although this heterogeneity differed in magnitude and significance through the three years of the study. Anderson and Karel [6] indicated dramatic haplotype frequency differences between SL and all other samples, and this is reflected in the current microsatellite data. Unfortunately, GB was not sampled in the current study. These two northernmost estuaries (GB and SL) are the remaining two estuaries into which TPWD practices marine stock enhancement. Similar to lower coast stocking, stocking practices in the upper coast dictate that fingerlings are released into the estuary from which brood stock are extracted, with consideration for logistical issues mentioned above. The current data support stocking in the bay of origin or into adjacent bays. Although it is impossible to evaluate differences between GB and SL with the current data specifically, caution must be urged for two reasons. First, the spatial abundance data presented here indicate that the ICWW is the only avenue of dispersal between these two estuaries, and it is unclear how frequently spotted seatrout use these areas as dispersal corridors, with any frequency. Second, the low intrapopulation haplotype diversity recovered in Anderson and Karel [6] may indicate that SL is demographically independent from sites further south. This clearly represents an avenue where future study is warranted.

# **Competing Interests**

This work was funded by two National Marine Fisheries Service Grants, through the Federal Aid in Sportfish Restoration program. Otherwise, the authors declare that they do not have any competing interests.

# **Authors' Contributions**

JA carried out all GIS tasks, mtDNA sequencing assays, statistical manipulations of both data sets, drafted the manuscript and made corrections. WK carried out all microsatellite DNA assays and assisted with analysis of microsatellite data.

#### Acknowledgement

We wish to thank the personnel of the Texas Parks and Wildlife Routine monitoring program for providing the samples used in genetic analyses. Heather Hendrickson assisted with laboratory tasks. Ashley Summers provided technical assistance with GIS analysis. Mark Fisher, Dusty McDonald and Britt Bumguardner as well as three anonymous reviewers improved the initial draft of the manuscript with helpful comments. TPW routine monitoring is carried out with the support of a NMFS Federal Aid in Sportfish Restoration Grant, F-34M. This research was specifically funded by NMFS Federal Aid in Sportfish Restoration Grant F-143-R.

#### References

[1] Green LM, Campbell RP, 2005. Trends in finfish landings of sport-boat anglers in Texas marine waters, May 1974-May2003. Texas Parks and Wildlife Management Data Series, No. 234.

[2] Weinstein MP, Yerger RW, 1976. Electrophoretic investigation of subpopulations of the spotted seatrout, Cynoscion nebulosus (Cuvier) in the Gulf of Mexico and Atlantic coast of Florida. Comparative Biochemistry and Physiology, 54B:97-102.

[3] Gold JR, Richardson LR, Furman C, 1999. Mitochondrial DNA diversity and population structure of spotted seatrout (*Cynoscion nebulosus*) in coastal waters of the southeastern United States. Gulf of Mexico Science, 1999:40-50.

[4] Ramsey PR, Wakeman JM, 1987. Population structure of *Sciaenops ocellatus* and *Cynoscion nebulosus* (Pisces: Sciaenidae): Biochemical variation, genetic subdivision and dispersal. Copeia, 1987:682-695.

[5] King, TL, Pate HO, 1992. Population structure of spotted seatrout inhabiting the Texas Gulf coast: an allozymic perspective. Transactions of the American Fisheries Society, 121:746-756.

[6] Anderson, JD, Karel WJ, 2009. A Genetic Assessment of Current Management Strategies for Spotted Seatrout in Texas. Marine and Coastal Fisheries, 1:121-132.

[7] Gold JR, Stewart LB, Ward R, 2003. Population structure of spotted seatrout (*Cynoscion nebulosus*) along the Texas Gulf Coast, as revealed by genetic analysis. In *Biology of the spotted seatrout*. Edited by Bartone SA. Boca Raton, FL: CRC Press, 17-29.

[8] Ward R, Bowers K, Hensley R, Mobley B, Belouski E, 2007. Genetic variability in spotted seatrout (*Cynoscion nebulosus*) determined with microsatellite DNA markers. Fishery Bulletin, 105:197-206.

[9] Excoffier L, Smouse PE, Quattro JM, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, 131:479-491.

[10] Excoffier L, Laval G, Schneider S, 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, 1:47-50.

[11] Nei M, Li W-H, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences, USA, 76: 5269-5273.

#### 19 Fisheries and Aquaculture Journal, Volume 2010: FAJ-2

[12] Ward JH, 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58:236-244.

[13] McGarigal K, Cushman S, and Stafford S, 2000. Multivariate Statistics for Wildlife and Ecology Research. New York:Springer-Verlag.

[14] Felsenstein J, 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics, 5:164-166.

[15] Turner TF, Richardson LR, Gold JR, 1998. Polymorphic microsatellite DNA markers in red drum (*Sciaenops ocellatus*). Molecular Ecology, 7:1771-1788.

[16] Saillant E, Cizdziel K, O'Malley KF, Turner TF, Pruett CL, Gold JR, 2004. Microsatellite markers for red drum, *Sciaenops ocellatus*. Gulf of Mexico Science, 22(1):101-107.

[17] Goudet J, 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

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

[19] Weir BS, Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution, 38:1358–1370.

[20] Cavalli-Sforza LL, Edwards AWF, 1967. Phylogenetic analysis: models and estimation procedures. American Journal of Human Genetics, 19:233-257.

[21] Anderson JD, McDonald DL, Sutton GR, Karel WJ, 2009. Evolutionary associations between sand seatrout (*Cynoscion arenarius*) and silver seatrout (*C. nothus*) inferred from morphological characters, mitochondrial DNA, and microsatellite markers. Fishery Bulletin, 107:13-23.

[23] Baker WB, Matlock GC, 1993. Movement of spotted seatrout tagged in Trinity Bay, Texas. Northeast Gulf Science, 13:29-34.

[24] Hedrick PW, 1999. Highly variable loci and their interpretation in evolution and conservation. Evolution, 53:313-318.

[25] O'Reilly PT, Canino MF, Bailey KM, Bentzen P, 2004. Inverse relationship between Fst and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. Molecular Ecology, 13:1799-1814.

[26] Balloux F, Brünner H, Lugon-Moulin N, Hausser J, Goudet J, 2000. Microsatellites can be misleading: and empirical and simulation study. Evolution, 54:1414-1422.

[27] Laikre L, Palm S, Ryman N, 2005. Genetic population structure of fishes: implications for coastal zone management. Ambio, 34:111-119.

[28] King TL, Zimmerman EG, 1993. Clinal variation at the aspartate minotransferase-2 in spotted seatrout, *Cynoscion nebulosus* (Cuvier), inhabiting the north-western Gulf of Mexico. Animal Genetics, 25:59-61.