Final Report: Analyzing genetic connectivity of Caribbean and Florida mutton snapper. By Kathryn Shulzitski, Dr. Michael McCartney, Michael L. Burton.

Introduction

Riley's Hump, in the recently created South Tortugas Ecological Reserve, is the historical site of a spawning aggregation of mutton snapper (*Lutjanus analis*), perhaps the only site in the southeastern United States. The aggregation was exploited heavily by commercial fishermen for decades prior to the area being declared an ecological reserve in 2001. Fishery managers hope that the population of mutton snapper will rebound now that the area is closed to all uses, including fishing.

In Year 1of a related study, we established stations for monitoring abundance of snappers and groupers to provide baseline estimates of abundance for comparison with future years' estimates in order to establish trends in populations. Protection from fishing pressure should allow the populations to increase, leading to "spillover", or increased recruitment of these fishes to areas outside the boundaries of the reserves. An important question to ask is: is the population at Riley's Hump the source of all recruits to the rest of the southeastern U. S? Previous detailed studies of the region's physical oceanography (Lee et al. 1994) show that the potential exists for the Tortugas Reserve to supply larvae to nearby waters. The effects of the Florida Current, onshore flow of surface waters, and the cyclonic circulation in the Tortugas gyre should combine to aid local retention and recruitment of larvae into the Florida Keys. Additional drifter studies show that larvae spawned in the Dry Tortugas may end up along the southeastern coast of Florida after entrainment in the Gulf Stream (Domeier 2004, Hare[1]). We need to address an equally relevant question: Where do the adult fish on Riley's Hump come from? Is the adult population at Riley's Hump dependent solely on itself to replenish the population, or is significant recruitment coming in from other upstream areas? In addition to the historical spawning aggregation at Riley's Hump, there are known aggregations further upstream in Belize (Domeier and Colin, 1997). We also know anecdotally from fishermen in Honduras (from the islands of Roatan and Utila) that aggregations of mutton snapper have been exploited there during the spawning season for a number of years. Both of these locations could conceivably be a source of recruits to the Riley's Hump/Tortugas population of mutton snapper, given oceanic circulation patterns in the region (Cowen et al. 2002).

Proper management of a species, especially one exploited as heavily as mutton snapper have been in the past, requires knowing the source of recruits to the fishery, and what kind of pressure these source populations experience. In this report we present the results of our ongoing FY03 work examining the connectivity between the Dry Tortugas and upstream sources such as the Mesoamerican Barrier Reef system off Belize and Honduras, as well as the connectivity of the Dry Tortugas with downstream areas along the Florida Keys and southeast Florida. The expected outcomes of this project were **1)** knowledge about the effectiveness of marine protected areas in rebuilding exploited fish stocks, and **2)** the development of scientific information about recruitment sources and sinks for U. S. populations of mutton snapper. This information should be useful to management agencies such as the South Atlantic Fishery Management Council (SAFMC)

[1] Hare, J. 2002. Personal commun. National Marine Fisheries Service, Naragansett, R. I.

and the FKNMS. This project directly addressed the identified priority areas of overfishing and threats to recruitment.

Methods

Samples

Mutton snapper from six geographic location were collected, and five locations were used in genetic analyses (Fig. 1). We were able to collect otoliths and tissue samples from approximately 50 mutton snapper each from spawning aggregations from Roatan (Honduras) and Belize in April 2003. We sampled the Roatan fish at a fish house in North Miami after they were imported by a local seafood dealer. The Belize fish were sampled for us by collaborating personnel of The Nature Conservancy in Belize. By sampling ourselves and contracting with fishermen, we collected samples from 40 fish from the Dry Tortugas by spring 2004. A local commercial fisherman and marine life collector was contracted to collect tissue samples from adult mutton snapper from the lower Florida Keys reef tract, mostly from one location near Ramrod Key, collecting 40 specimens by summer 2004. These samples from the lower Keys were archived, but have not yet been genotyped. We used a NMFS port agent to collect tissue samples from 50 adult mutton snapper landed by the southeast Florida headboat fishery in the summer of 2004. However, these fish were unfortunately lost when Hurricane Frances hit

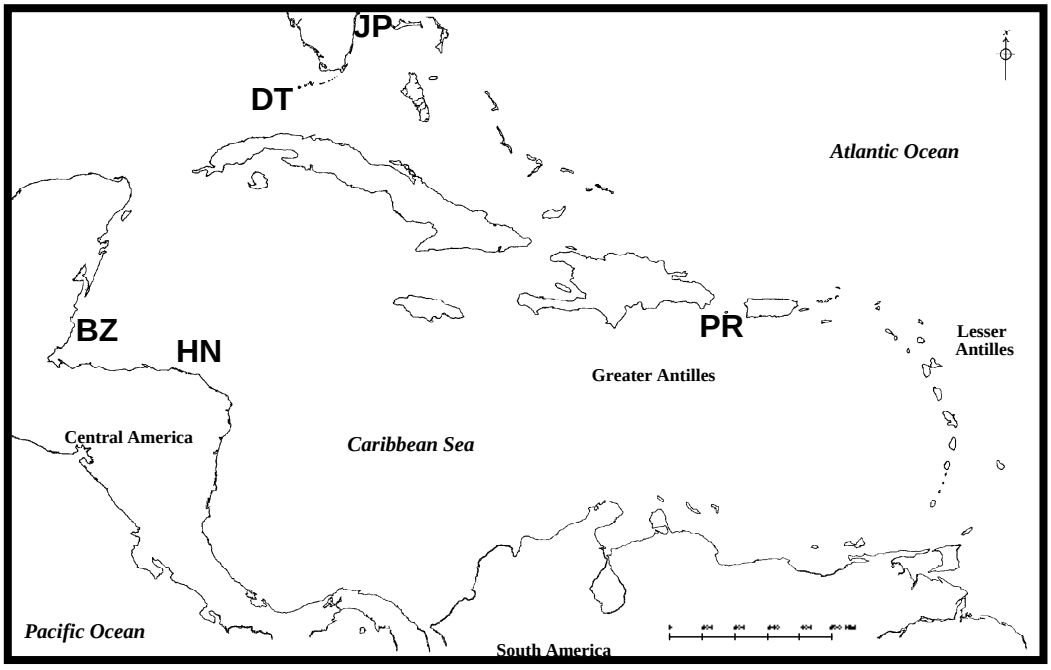

Figure 1. Map of sampling locations. Location codes, with sampling dates and sample sizes (N) from each location, are: BZ = Gladden Spit, Belize, May 2003, N = 50; DT = Dry Tortugas, June 2003 to December 2003, N = 40; HN = Roatan, Honduras, April 2004, N = 53; JP = Jupiter, FL, October 2004, N = 55; PR = western Puerto Rico, September 2004 to February 2005, N = 47.

southeast Florida in September, and extended power outages and excessive heat led to deterioration of our samples. We compensated for this loss by collecting tissue samples from juvenile mutton snapper in the estuary, using fishery independent seine sampling, in October, successfully procuring 50 samples over a two day sampling event, with the help of biologists from the Florida Wildlife Research Institute's Tequesta fisheries lab. Additionally, we contracted with a local fisherman in western Puerto Rico for collection of 50 specimens from the commercial fishery there, in order to obtain some eastern Caribbean specimens. These fish were collected in fall and winter of 2004. We planned on obtaining samples from at least one more eastern Caribbean location (e.g., Turks and Caicos, Barbados, Trinidad and Tobago, the USVI, Cuba) but we were unable to establish successful contacts at any of these locations.

Development of microsatellite loci

The first portion of the molecular work on this project required the isolation and characterization of microsatellite DNA loci from mutton snapper, since none were previously available. Microsatellite-enriched genomic DNA libraries were created. Genomic DNA of four adult fish from Honduras was isolated and enriched for microsatellites by the Savannah River Ecology Lab in Aiken, South Carolina according to the procedures of Glenn et al. (2000). We received the enrichment products and amplified them by the polymerase chain reaction (PCR). PCR products were purified and cloned into pGEM-T Easy vector (Promega, Madison, WI). We selected a total of 200 clones, 100 from each of 2 libraries, and sequenced these clones using Big Dye cycle sequencing (Applied Biosystems) and analyzed the sequences on the ABI 3100 automated Genetic Analyzer at UNCW.

Sequences were screened for microsatellites (i.e. tandem repeats from two to six base pairs in length with a minimum of six repeat units) in *Sequencher 4.2* (GeneCodes) and classified according to Weber (1990). Clones containing a microsatellite repeat were targeted for PCR primer design in the web-based program *Primers 3* (Rosen and Skaletsky 2000). One primer in each primer pair was "PIG-tailed" (Brownstein et al. 1996) in order to force the A-tailing of all PCR products. This substantially reduced the ambiguity in allele sizing that can result from the tendency of *Taq* polymerase to add an A-tail to some products and not to others.

Microsatellite DNA genotyping

In initial genotyping trials, a total of ten individuals each from Honduras, Belize and the Dry Tortugas were genotyped on the ABI 3100 using the G5 dye system, which allows the products of four microsatellite loci from each individual to be analyzed simultaneously in a single capillary. The forward primer (not PIG-tailed) for each locus was fluorescently-labeled with PET, NED, 6FAM and VIC dyes. Genotypes were analyzed against the size standard with GENESCAN 3.7 (Applied Biosystems) and assigned to size categories in GENOTYPER 3.7 (Applied Biosystems). Genotypes of the thirty individuals were tested for departure from Hardy-Weinberg equilibrium (HWE) at each locus in *GENEPOP* (Raymond and Rousset 1995a) using an exact probability test and a Markov chain method to obtain the unbiased exact P-value (Guo and Thompson 1992). A highly significant departure from HWE, calculated separately for each

sampling location, was considered potential evidence for the presence of null alleles and thus that locus was dropped from the study or, in two cases, primers were redesigned. The annealing temperature for each remaining set of primers was optimized using an MJ Research thermal cycler with a gradient block. Genomic DNA from three individuals exhibiting a range in genotype quality (e.g. presence of non-specific amplification) was amplified with annealing temperatures ranging from 55 to 65°C. Primer pairs that produced unambiguous peaks, appropriate levels of polymorphism, and conformed well to the expectations of Mendelian markers (i.e. one or two peaks present in each individual) and HWE were chosen for large-scale genotyping.

For large-scale analysis, approximately 245 individuals total were genotyped at eight loci (Tables 1. 2). Amplifications were 15 µl in volume and contained 5-20 ng template DNA, 15 mM Tris-HCl, pH 8.0, 50 mM KCl, 1.5 mM MgCl2, 2.5 mM each dNTP, 0.5 uM each primer, and 0.75 U of AmpliTaq Gold (Applied Biosystems). Cycling parameters began with a hot start of 10 minutes at 95°C, followed by 35 cycles of 30 seconds at 94°C, 30 seconds at the designated annealing temperature (see Table 7), and 60 seconds at 72°C, with a final extension of 30 minutes at 72°C. PCR products were visualized on an ABI 3100 and analyzed with *GENESCAN 3.7*. In *GENOTYPER 3.7*, peaks were labeled and binned into size categories corresponding to allelic length in base pairs.

Data Analysis

Each population was tested for departures from HWE at each locus in *GENEPOP* (Raymond and Rousset 1995a) using a probability test and a Markov chain method to obtain the unbiased exact *P*-value (Guo and Thompson 1992). In a similar manner, all locus pairs were tested for linkage equilibrium within each population and across all populations. Since sample sizes varied from 38 to 55 individuals, allelic richness, the number of alleles present in populations independent of sample size, was calculated for each population-locus combination as well as overall using *FSTAT* (Goudet 1995). This calculation estimated the expected number of alleles for a sub-sample of genes equal in size to that of the smallest sample.

Population-based analyses were utilized to test the null hypothesis that all five mutton snapper samples comprise a genetically homogenous, panmictic population. First, in order to assess allelic and genotypic distributions across populations, an exact probability test (Raymond and Rousset 1995b) and a log-likelihood based exact test (Goudet et al. 1996) were performed in *GENEPOP* to evaluate genic and genotypic differentiation, respectively. Both tests employed a Markov chain method to calculate an unbiased estimate of the *P*-value.

A second evaluation, Wright's (1921) fixation index, is based on heterozygote deficiencies that result from non-random mating. As a measure of among-population subdivision, the unbiased estimator of F_{ST}, θ (Weir and Cockerham 1984), was calculated in *GENETIX* (Belkhir et al. 1996-2002), as were pairwise F_{ST} values, and permutation tests (1000 randomizations) were used to generate *P*-values. An estimator of R_{ST}, ρ (Slatkin 1995), was also calculated using *FSTAT* (Goudet 1995). This analog of F_{ST} takes into account allelic size by assuming alleles of a similar size are more closely related, given that loci adhere to a stepwise mutation model (Slatkin 1995). An analysis of molecular variance (AMOVA, Excoffier et al. 1992) based on the number of different

alleles (F_{ST}) was performed in *ARLEQUIN* (Schneider et al. 2000). This hierarchical analysis distributes the observed variance in heterozygosity into within population, between population and between group components. Finally, genetic chord distance, D_{CE} (Cavalli-Sforza and Edwards 1967), between each population pair was calculated in *GENETIX* with permutation tests (1000 randomizations) to estimate *P*-values.

Individual-based analyses were also used to assess population structure in the five mutton snapper samples. These analyses exploit more of the information contained in each individual multilocus genotype, in contrast to population-based analyses that primarily utilize allele frequencies and heterozygosities calculated for each population. Two fundamentally different procedures were used to assign individuals to source populations based on their multilocus genotypes: frequency-based and Bayesian assignment. The frequency-based method of Paetkau et al. (1995) was implemented in *GENECLASS* (Piry et al. 2004). Populations were determined *a priori* based on sampling locations, and *GENECLASS* generated allele frequencies for each population excluding the individual to be assigned (Waser and Strobeck 1998). The expected frequency of each individual's genotype at each locus across all populations was calculated and each individual was assigned to the population from which its multilocus genotype most likely originated. A frequency of 0.001 was designated to alleles that were absent from a population. In order to generate probabilities rather than likelihood values, *GENECLASS* performs a Markov Chain Monte Carlo (MCMC) resampling procedure with 10,000 simulated individuals (Paetkau et al. 2004).

Genotypes were also analyzed using a Bayesian population assignment procedure implemented in the program *STRUCTURE* (Pritchard et al. 2000). *STRUCTURE* assumes Hardy-Weinberg and linkage equilibrium and uses a MCMC algorithm to infer population parameters that conform to these modeling assumptions. In order to determine the number of genetically distinct clusters (K) in the dataset, individual genotypes were used as the prior in the estimation of posterior probabilities for K = {1, 2,…5}. Five runs were performed at each value of K to ensure proper mixing in the chain and consistent results. All runs used a burnin of 10^6 followed by 10^6 MCMC iterations. Parameters assumed an admixture model in which individuals may have mixed ancestry, and correlated allele frequencies which could account for similarity between closely related populations. For each individual in the analysis, a probability of membership in each cluster was estimated.

Results

Development of microsatellite loci

A total of 76 microsatellite repeat sequences were isolated from the enriched genomic libraries. These included 53 di-, three tri-, 18 tetra- and two pentanucleotide repeats. Based mostly on ease of PCR amplification and structure of the repeat, 18 loci were chosen for dye-labeling of the forward primer (Table 1). All eighteen sets of dye-labeled primers successfully amplified 10-30 individuals at target loci. Based on repeat motif, number of alleles, observed heterozygosity, conformation to Hardy-Weinberg expectations (Table 1), eight of these loci were chosen for large-scale genotyping (Table 2). Several dinucleotide loci showed unacceptable levels of stutter banding and were

discarded, and the chosen loci produced high quality unambiguous genotypes (Figure 2). Primer sets consisting of four loci each allow for the most efficient utilization of the ABI 3100 since four dyes can be simultaneously loaded into each capillary. The loci chosen for final application include two di- and six tetranucleotides. These loci exhibited a range in polymorphism with 12 to 36 alleles at a locus and observed heterozygosities from 0.700 and 1.000 (Table 2). Despite the presence of significant deviations from HWE at La27a, this locus was chosen based on the repeat motif, relatively high observed heterozygosity, and the ease of allele identification and sizing.

Table 1. Loci targeted for fluorescent dye-labeling. N = number of individuals genotyped per locus; H_E = expected heterozygosity; H_O = observed heterozygosity; test for HWE deviation: *** $P < 0.001$.

Locus	Repeat motif of clone	N	Number of Alleles	H_E	H_O	Comments
LaA-4	$(AAG)_{13}$					monomorphic
LaA-12	$(CA)_{14}$	30	7	0.626	0.600	
LaC-16	$(TC)_{11}(AC)_8$	33	10	0.890	0.939	
La8	$(CA)_{17}GA(CA)_{21}$	31	18	0.818	0.806***	null alleles
La13	$(AG)_{24}CG(AG)_{21}$					stutter bands
La25	$(TAGA)_{19}$	32	21	0.939	0.906	
La34	$(AC)_{18}$	32	23	0.946	0.969	
La39	$(CA)_{11}$	32	13	0.835	0.906	
La48	$(AC)_{18}$					stutter bands
La18a	$(GGAT)_5GGGT(GGAT)_2(AGAT)_{21}$	32	36	0.976	1.000	
La19a	$(GATA)_{12}$	29	12	0.848	0.345***	null alleles
La20a	$(AGAT)_5AGGTAGAT(AGGT)_{16}$					non-specific amplification
La23a	$(AGAT)_{13}$	41	13	0.868	0.634***	null alleles
La27a	$(AGAT)_{21}$	20	16	0.915	0.700***	null alleles
La27a ‡		20	17	0.923	0.700***	
La31a	$(AGAT)_{18}$	28	15	0.894	0.536***	null alleles
La36a	$(TAGA)_{26}$	10	5	0.784	0.200***	null alleles
La36a ‡		10	7	0.789	0.400***	
La45a	$(CCAT)_{15}$	26	12	0.883	0.962	
La49a	$(AGAT)_{11}$	22	12	0.910	0.955	

‡Denotes the redesigned primer sets.

Significant deviations from HWE were present at six loci (including La27a) in the form of heterozygote deficiencies. Several explanations for this departure from HWE exist, including population substructure and the presence of null alleles. Although tests were conducted separately for each location, it is not unexpected to find population substructure within a geographic sample simply due to chance events. Yet null alleles, alleles that fail to amplify due to mutations within priming sites, are reportedly common across microsatellite loci (Callen et al. 1993) and can occur in relatively high frequencies (Paetkau and Strobeck 1995). In addition, the presence of null alleles could significantly affect the results of a study. Therefore, in order to err on the side of caution, departures from HWE within a single location were considered evidence of null alleles.

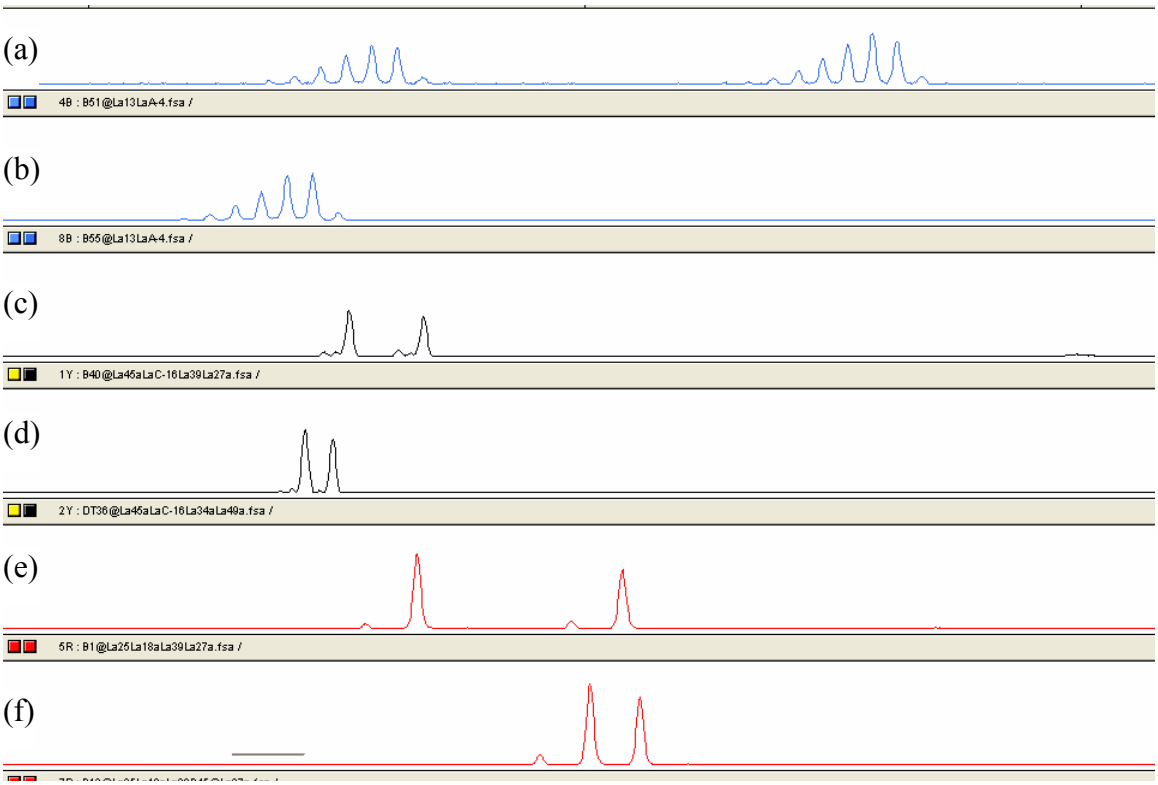

Figure 2. Stutter banding at locus La13 compared to peak morphologies at loci chosen for large-scale genotyping. (a) represents a heterozygote for two alleles, the size of both of which was ambiguous, (b) a potential homozygote, (c) and (d) heterozygotes at LaC-16, a dinucleotide locus, and (e) and (f) heterozygotes at La25, a tetranucleotide locus.

To address this issue further, two loci, La27a and La36a, were chosen for primer redesign (Table 1). A small sample of previously typed individuals was genotyped using a new set of primers for La27a and a new forward primer for La36a. At locus La27a, one previously homozygous individual was typed as a heterozygote with redesigned primers and one previously heterozygous individual was typed as a homozygote. At locus La36a, two previously homozygous individuals were typed as heterozygotes with redesigned primers. Although the sample sizes employed were not large, this result establishes the presence of null alleles for at least two loci in this study. The frequency of nulls was estimated to be 0.116 and 0.330 at loci La27a and La36a, respectively (Brookfield et al. 1996).

Table 2. Mutton snapper microsatellite loci. T_m = annealing temperature; N = number of individuals genotyped; A = number of alleles; H_E = expected heterozygosity; H_O = observed heterozygosity. Test for Hardy-Weinberg equilibrium: * $P < 0.05$, *** $P < 0.001$

Locus	Repeat motif of clone	Primer sequences (5' to 3')	Tm (°C)	Product size range (bp)	N	A	H_E	H_O
La25	$(ATCT)_{19}$	**F:** GGA GGA ACC TCC TGG AAT GT **R:** GTT TGC ACT TGA AGA AAA AGG GTG A	63	137-235	241	36	0.942	0.926
La39	$(TG)_{11}$	**F:** TGC TGA GGA GCA TTT GCT TT **R:** GTT TAA AGT CAC ATA AAC GGG GAC T	65	143-179	226	15	0.802	0.603***
La18a	$(ATGG)_8GTGGATGG(ATAG)_{21}$	**F:** CCT CAC TTT TGT GTG AGA CAG C **R:** GTT ATG ATT GCT AGG AGC ATC TGG	65	188-340	239	37	0.964	0.963
La27a	$(AGAT)_{22}$	**F:** CTT AGC AAG CCA ACA AAC AAT G **R:** GTT TCC AAG GTC CAT TGA TCT TTA GTG	65	176-312	184	39	0.939	0.856*
La34a	$(AGAT)_{14}(AGTG)(GAGT)(GAGA)$	**F:** TGT CTC TTC GAA ATC AAA CAC AA **R:** GTT TGA GGC TTA TCT GCC CCT CTC	57	239-303	235	17	0.914	0.908
La45a	$(TCCA)_{15}$	**F:** AAC CAC ATC TGG CTC AAT CA **R:** GTT TAG CCC CAG AGT AGG GTG AGA	62	194-254	238	15	0.880	0.871
La49a	$(TATC)_{12}(CATC)(TATC)$	**F:** GCT GAG GCA GAA ATC ACA C **R:** GTT ATG TCC ACT GAT GCC TCA AAA	65	224-312	235	19	0.905	0.840
LaC-16	$(TC)_{11}(AC)_8$	**F:** GGT GTT GAT TGG TCC TCT GG **R:** GTT TGG GGT TGG TAT TCA TCC AGT	66	146-164	238	10	0.853	0.834

Large-scale genotyping results

High levels of polymorphism were observed in all five populations of mutton snapper at the eight microsatellite loci. The expected and observed heterozygosities ranged from 0.794 to 0.966 and 0.500 to 0.982, respectively (Table 3). The number of alleles detected per locus ranged from nine to 32. As can be expected, populations with larger sample sizes exhibited slightly increased levels of allelic diversity, with a total of 149 alleles present in JP and only 139 in DT. Also, only two private alleles were present in DT, while all other populations contained seven or eight. However, there were no apparent trends towards reduced heterozygosity in populations with smaller sample sizes, and estimations of allelic richness indicated that no single population was particularly deficient in genetic diversity.

Seven out of 40 tests indicated significant departures from HWE ($0 < P < 0.05$). Four of these significant tests occurred at locus La39, while the remaining three were distributed across loci (Table 3). PR and BZ each exhibited significant deviations from HWE at two of the eight loci. All locus pairs exhibited linkage equilibrium when tested across populations. In contrast, pairwise locus comparisons within each individual population yielded six significant tests indicating linkage disequilibrium out of a total of 140 tests ($0.01 < P < 0.05$). Interestingly, five of these significant tests occurred in the PR sample.

Population Structure

Three out of 80 tests indicated significant heterogeneity in allele frequencies between population pairs ($0.01 < P < 0.05$). Each of these three tests included PR as compared to JP, BZ and HN, and heterogeneity was limited to La18a and La39 (Fig. 3). Tests of genotypic distributions between population pairs indicated significant heterogeneity in two out of 80 tests ($0.01 < P < 0.05$).

Table 3. Genetic diversity at eight microsatellite loci in five populations of mutton snapper. N = number of genotyped individuals; A = number of alleles; Â = allelic richness; a = number of private alleles; H_E = expected heterozygosity; H_O = observed heterozygosity; test for HWE: * $P < 0.05$, ** $P < 0.01$, *** $P \leq 0.001$.

	Locus								
Population	La25	La39	La18a	La27a	La34a	La45a	La49a	LaC-16	Overall
Belize									
N	49	44	47	39	48	49	48	48	
A	21	11	31	25	14	13	14	10	139
Â	18.9	9.6	27.0	22.3	13.4	11.5	13.0	9.4	125.1
a	-	-	3	2	-	-	2	-	7
H_E	0.943	0.806	0.968	0.945	0.910	0.856	0.911	0.864	
H_O	0.939	0.750	0.957	0.872*	0.875	0.898	0.896	0.896*	
Honduras									
N	53	53	53	40	52	52	52	53	
A	27	13	28	26	17	14	12	10	147
Â	21.5	11.3	24.5	22.3	15.6	12.3	11.5	9.5	128.5
a	5	1	-	2	-	-	-	-	8
H_E	0.942	0.813	0.962	0.939	0.926	0.887	0.905	0.847	
H_O	0.868	0.585***	0.962	0.825	0.942	0.865	0.827	0.830	
Dry Tortugas									
N	38	38	38	29	38	39	38	39	
A	24	11	31	19	16	13	12	9	135
Â	21.8	10.3	28.1	19.0	15.4	12.0	11.3	8.7	126.6
a	-	1	-	1	-	-	-	-	2
H_E	0.942	0.794	0.964	0.920	0.915	0.876	0.894	0.859	
H_O	0.895	0.500***	0.947	0.862	0.921	0.821	0.789	0.846	
Jupiter									
N	55	54	55	44	55	55	55	54	
A	25	11	32	29	16	12	14	10	149
Â	21.2	9.7	26.8	23.9	14.8	11.4	12.8	9.3	129.9
a	2	-	2	2	-	-	1	-	7
H_E	0.949	0.802	0.966	0.945	0.917	0.891	0.916	0.858	
H_O	0.964	0.537***	0.982	0.841	0.836	0.873	0.891	0.759	
Puerto Rico									
N	46	37	46	32	42	43	42	44	
A	25	12	27	22	16	12	15	10	139
Â	21.0	10.6	23.7	21.0	14.9	11.2	13.9	9.3	125.6
a	3	-	-	1	-	1	2	-	7
H_E	0.934	0.771	0.956	0.937	0.911	0.873	0.908	0.846	
H_O	0.957	0.622**	0.957	0.906	0.976	0.884	0.762*	0.841	

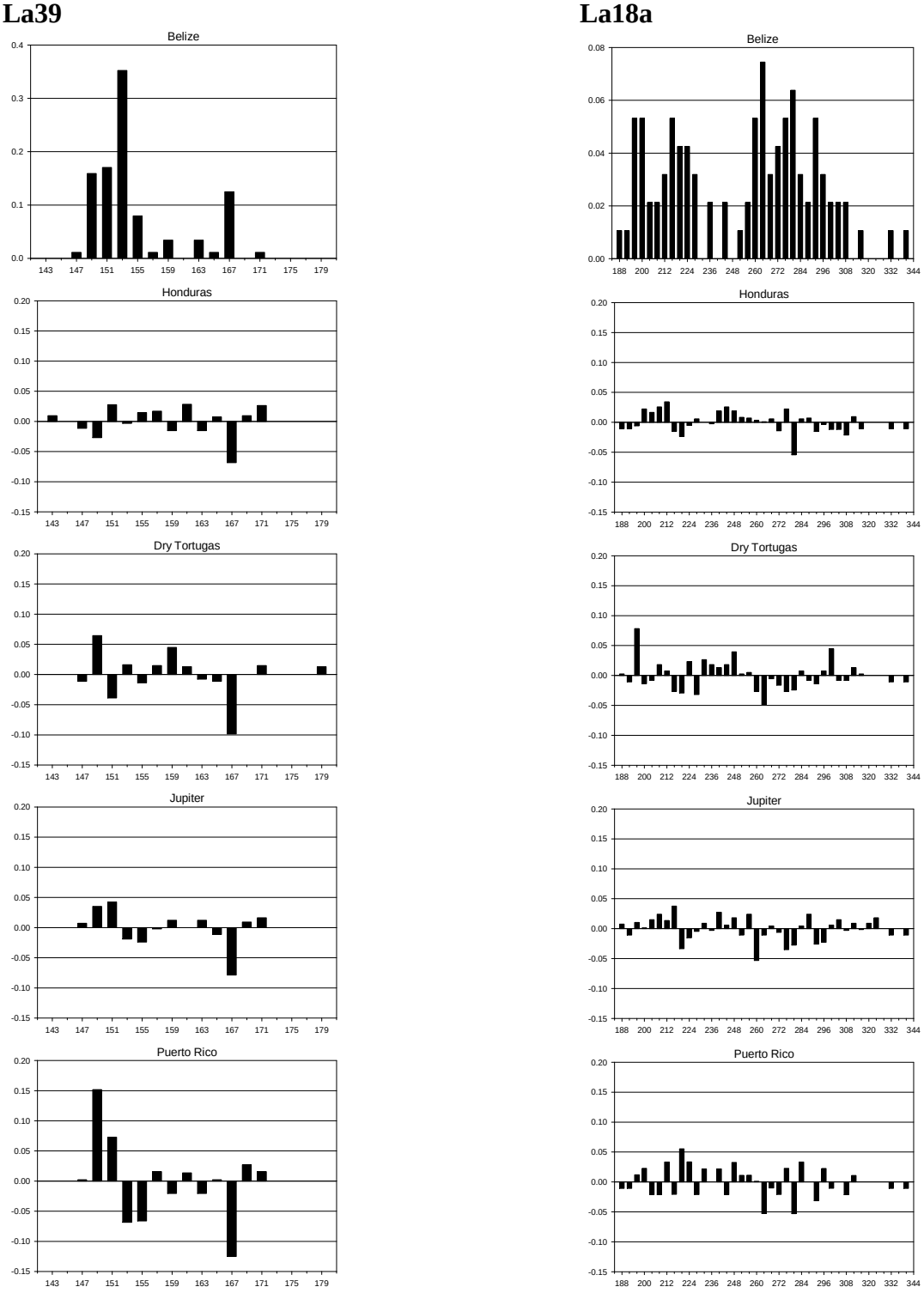

Figure 3. Allele frequencies at two microsatellite loci in five mutton snapper populations. The absolute allele frequencies are shown for BZ, while the other four populations are shown as the difference with the observed distribution in BZ. For each locus: x-axis = allele size (bp) and y-axis = allele frequencies.

Table 4. Population structure across all populations estimated with θ (F_{ST}) and ρ (R_{ST}).

Locus	θ	ρ
La25	0.000	-0.003
La39	0.000	0.005
La18a	0.000	-0.005
La27a	-0.001	0.000
La34a	-0.002	-0.008
La45a	0.003	-0.003
La49a	-0.005	-0.009
LaC-16	-0.004	-0.009
Overall	-0.0011	-0.0041

Table 5. Pairwise comparisons of mutton snapper populations. Above the diagonal: F_{ST} (*P*-value); below the diagonal: D_{CE} (*P*-value).

	BZ	HN	DT	JP	PR
BZ		-0.0024 (0.926)	-0.0015 (0.748)	-0.0016 (0.844)	0.0007 (0.312)
HN	0.012 (0.921)		0.0004 (0.401)	-0.0035 (0.995)	-0.0009 (0.661)
DT	0.014 (0.842)	0.015 (0.686)		-0.0023 (0.878)	0.0022 (0.176)
JP	0.012 (0.792)	0.012 (0.900)	0.014 (0.913)		-0.0001 (0.468)
PR	0.017 (0.119)	0.014 (0.747)	0.018 (0.405)	0.017 (0.096)	

Table 6. Analysis of molecular variance (AMOVA) among groups and populations. All tests of F-statistics were not significant.

Source of variation (FST)	Degrees of freedom	Sum of squares	Variance components	Variation (%)
Among groups	1	1.714	-0.00341	-0.13
Among populations within groups	3	6.661	-0.00357	-0.14
Within populations	485	1247.596	2.57236	100.27
Total	489	1255.971	2.56538	

All estimates of θ (F_{ST}) and ρ (R_{ST}) across populations were less than 0.005 and most were zero or negative values (Table 4). Pairwise comparisons of F_{ST} ranged from -0.0035 to 0.0022, and all values were not significant. There was a slight trend in which pairwise comparisons involving PR exhibited the largest, albeit still small, F_{ST} values; the DT with HN comparison was the exception to this trend (Table 5). An AMOVA was performed with PR pre-defined as one group and the remaining populations comprising a second group. The definition of groups was based on the results of previous work that indicated Puerto Rico as a potential intermediate population between eastern and western Caribbean genetic clusters (Baums et al. 2005). One-hundred percent of the variance was partitioned to the within populations component (Table 6). The same result was obtained when the analysis was performed based on the sum of squared size differences (R_{ST}) and when HN was included in the first group with PR (data not shown). Pairwise genetic distances, D_{CE}, were all non-significant and ranged from, 0.012 to 0.018. Again, the largest distances were calculated for comparisons involving PR. The only discrepancy in this trend was the DT with HN comparison. Although the results of these population-level analyses were not significant, the tendency of the Puerto Rican population to show the largest differences between samples supports distinction of this geographic location.

Population Assignment

Only 35 individuals (14.3%) were correctly assigned to their source populations (Table 7). Individuals were assigned to the population in which their multilocus genotypes exhibited the highest probability of occurrence. However, assignment probabilities in the data set ranged from 0.0011 to 0.9416 and, for most individuals, these values were similar across populations. Thus, this method was unsuccessful at detecting the origin of an individual based on its multilocus genotype and population allele frequencies. This is not surprising given the similarity in allele frequencies across population samples. In addition, sample size appeared to highly influence the assignment procedure, as the two largest populations, HN and JP, received the greatest number of assignments (Table 7).

STRUCTURE identified a single genetic cluster with a posterior probability of 1.0 (Table 8). Thus, in the absence of prior information regarding the origin of individual genotypes, posterior probabilities of the parameter, K, do not support the presence of genetic structure within the sample of 245 individuals. When the assignment procedure was implemented using five clusters, individuals from each population were assigned equally well to all five (Table 9). A similar result was obtained assuming two clusters for the assignment (Table 9). Again, this result is not particularly unexpected based on the results of previous tests indicating that the populations conformed well to expectations of Hardy-Weinberg and linkage equilibrium.

Table 7. Individual assignments of the frequency-based test. Values represent the number (%) of individuals from the source populations in each assigned population. Bold-faced values denote assignment to populations of origin.

	Assigned Population					
Source Population	BZ	DT	HN	JP	PR	Total
BZ	**3 (6.0)**	7 (14.0)	16 (32.0)	23 (46.0)	1 (2.0)	50 (20.4)
DT	4 (10.0)	**5 (12.5)**	13 (32.5)	15 (37.5)	3 (7.5)	40 (16.3)
HN	6 (11.3)	11 (20.8)	**10 (18.9)**	21.5 (40.6)	4.5 (8.5)	53 (21.6)
JP	9 (16.4)	11 (20.0)	18 (32.7)	**14 (25.5)**	3 (5.5)	55 (22.4)
PR	5 (10.6)	6 (12.8)	19 (40.4)	14 (29.8)	**3 (6.4)**	47 (19.2)
Total	27 (11.0)	40 (16.3)	76 (31.0)	87.5 (35.7)	14.5 (5.9)	245

Table 8. Probability of the number of populations (K) inferred by STRUCTURE. Log P(X|K) = ln probability of the data and log P (K|X) = posterior probability.

K	log P(XIK)	P(KIX)
1	-9631	1.0
2	-9914	~0.0
3	-10409	~0.0
4	-10544	~0.0
5	-11850	~0.0

Table 9. Proportion of membership of each pre-defined population in clusters inferred by STRUCTURE assuming two genetic clusters (a) and five genetic clusters (b).

Population ID	Two inferred clusters		Five inferred clusters				
	I	II	I	II	III	IV	V
BZ	0.500	0.500	0.204	0.199	0.198	0.200	0.200
HN	0.500	0.500	0.215	0.186	0.203	0.187	0.209
DT	0.500	0.500	0.192	0.208	0.198	0.207	0.195
JP	0.500	0.500	0.201	0.200	0.200	0.200	0.199
PR	0.500	0.500	0.198	0.207	0.195	0.204	0.197

Discussion

The null hypothesis, that sampled mutton snapper populations constitute a single panmictic population, cannot be rejected based on the genetic pattern observed at eight microsatellite loci across 245 individuals. Analyses ranging from population-based F-statistics to individual-based assignment indicate that population genetic substructure is absent from the five sample locations, ranging across approximately 2000 km.

A lack of genetic differentiation across large distances is not unique to mutton snapper populations. Heist and Gold (2000) obtained similar results with five microsatellite loci across 1500 km of the Gulf of Mexico for the red snapper, *Lutjanus campechanus*. Genetic diversity was slightly higher in the present study; however, values of F- and R-statistics were similar for both red and mutton snapper populations. Larval dispersal has long been recognized as a homogenizing force in marine systems. The pelagic larval duration (PLD) of mutton snapper is 27 to 37 days based on otolith settlement marks (Lindeman et al. 2000), similar to that of the red snapper (28 to 30 days, Heist and Gold 2000). This 4- to 5-week period of time spent in the water column may facilitate long distance dispersal and, thus, genetic homogenization.

In contrast to the outcome of the present study, microsatellite loci have often been recognized for their ability to reveal subtle genetic structure. As was previously discussed, populations of the gorgonian coral, *Pseudopterogorgia elisabethae*, exhibited significant structure across small geographic distances at six microsatellite loci (Gutierrez-Rodriguez and Lasker 2004). Yet it must be noted that the life history of these brooding corals is quite distinct from that of mutton snapper, which broadcast gametes in large offshore spawning aggregations, resulting in relatively long-lived pelagic larvae. Significant genetic structure was also detected using microsatellites in the elkhorn coral, *Acropora palmata* (Baums et al. 2005). Although these corals broadcast gametes into the water column, complex inshore currents coupled with a five- to 20-day PLD may facilitate local retention. Thus, based on life history characteristics alone, it is not surprising that the five mutton snapper populations exhibited genetic homogeneity. However, it is important to recognize several caveats to the conclusion of panmixia for mutton snapper populations based on the present study.

The range of mutton snapper across the Caribbean, in particular its large spawning aggregations, was not sampled in its entirety. According to the only other study employing microsatellite markers to a similar geographic range, genetic differentiation was partitioned into eastern and western Caribbean clusters (Baums et al. 2005). Samples of elkhorn coral from Puerto Rico and Mona Island were intermediate between the two clusters. Mutton snapper populations belonging to the eastern cluster were not sampled, so we cannot evaluate divergence between eastern and western Caribbean populations. Interestingly, however, there was a non-significant trend in which PR was the most distinct of the five populations. This is apparent in the involvement of PR in all significant tests for deviations from H-W and linkage equilibrium, and for heterogeneity in allelic and genotypic distributions. The trend is also present in pairwise estimates of F_{ST} and genetic chord distance. Due to the tendency of larval dispersal to homogenize populations, the longer larval duration of the mutton snapper relative to that of elkhorn coral may cause genetic structure to be more subtle in the former. This being the case, the slight and non-significant differentiation of PR may be consistent with the results of Baums et al. (2005). However, no definitive conclusions

can be made without the addition of eastern Caribbean populations to the mutton snapper dataset.

Although microsatellites have proven to exhibit high resolution in numerous studies, their application remains hindered by the disproportionate effect of very small levels of migration on genetic divergence (Slatkin 1987). Thus, the results of genetic surveys can greatly inflate estimations of dispersal between populations that are demographically closed on an ecological scale. This issue of scale is apparent in the conflicting results of studies employing alternative markers to measure local retention. For example, significant levels of natal homing were revealed for weakfish along the eastern coast of the United States (Thorrold et al. 2001). Sixty to 81 % of individuals returned to spawn in the estuary of their birth. However, previous genetic studies employing allozymes and mtDNA failed to detect any genetic structure in weakfish sampled across a similar range. Thus, although genetic markers exhibit homogeneity among mutton snapper populations, alternative markers (e.g. elemental signatures or artificial tags) may reveal significant levels of local retention around spawning sites.

Management and Conservation Implications

The genetic similarity of mutton snapper populations revealed at eight microsatellite loci implies free gene flow between populations. Hence, because the genetic composition of each sample was so similar, it is impossible to discern between the relative contributions of Puerto Rico, Belize, Honduras and the Dry Tortugas as source populations for the Jupiter sample. Thus we cannot rule out the possibility that the DTER serves as a significant source of recruits to the southeastern United States. Yet based on the results of this study we cannot confirm that is does. It is logical to assume that the nearest source population (i.e. Dry Tortugas) supplies the majority of recruits to southeast Florida; however, in the absence of better resolved genetic structure, this will remain unknown. Future attempts to estimate the relative contributions of mutton snapper spawning aggregations to adult populations may include an increase in the number of loci employed. This would be quite feasible given that one valuable product of this study is enriched genomic libraries that contain countless other microsatellite loci. An alternative strategy would involve a large increase in sample size. The addition of individuals could increase the resolution by revealing more rare alleles that are confined to particular locations. The development of alternative markers, such as single nucleotide polymorphisms (SNP) or nuclear sequences, is also possible, though much more expensive and time-consuming. Finally, the implementation of mark-recapture studies using artificial tags could reveal the degree of site fidelity exhibited by adults as well as the geographic range of adults that utilize a particular aggregation. This avenue of research would also be feasible, as numerous adults could be marked at once in the spawning aggregation.

Future Directions for Understanding Connectivity

The apparent genetic distinctiveness of the Puerto Rico population (Fig. 3) suggests that additional genetic surveys in mutton snapper, and in other carefully selected fisheries species, would be of great value. In particular, studies that survey both the Antilles and Mesoamerica; that is, both the eastern and western Caribbean, are an important next step. Such work should survey genetic variation at microsatellite loci for

a variety of species sampled from across the Caribbean basin. Shulman and Bermingham (1995) used mitochondrial DNA restriction fragment length polymorphism and found significant population structure in 3 reef fish species. Future work with newer high resolution markers should enable the detection of the more subtle population structure that can be expected for species with high dispersal potentials. The goal of such a study would be to understand the dominant dispersal corridors, most likely driven by the major surface current patterns, and hence to predict connectivity between potential MPA sites across the Caribbean. Species exhibiting a range of dispersal potentials, in particular species with minimal pelagic larval durations (PLDs), should also be incorporated into this work. The range of estimated PLDs surveyed in Shulman and Bermingham (1995) averaged from 15 to 81.5 days. In order to increase the potential for evaluating genetic connectivity, future work should include commercially valuable species but also fishes with very low dispersal powers (e.g. gobies). Such species could be used as "indicator species" to infer broad patterns of connectivity that are likely to also influence demographic connectivity between populations of commercially targeted species, like snappers and groupers.

References

Baums IB, Miller MW, Hellberg ME (2005) Regionally isolated populations of an imperiled Caribbean coral, *Acropora palmata*. *Molecular Ecology* **14**, 1377-1390.

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. 1996-2004 GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Brookfield, J. F. Y. 1996 A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**, 453-455.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**, 1004-1010.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics* **52**, 922-927.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics* **19**, 233-257.

Cowen, R.K., C. B. Paris, D. B. Olson, and J. L. Fortuna (2002) The role of long distance dispersal versus local retention in replenishing marine populations. *Gulf and Caribbean Research* 14, 129-137.

Domeier, M. L (2004. A potential larval recruitment pathway originating from a Florida marine protected area. *Fisheries Oceanography* 13, 287-294.

Domeier, M. L. and P. L. Colin (1997) Tropical reef fish spawning aggregations: defined and reviewed. *Bulletin of Marine Science* 60, 698-726.

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.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
Goudet J, Raymond M, de Meeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics* **144**, 1933-1940.
Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-372.
Gutierrez-Rodriguez C, Lasker HR (2004) Microsatellite variation reveals high levels of genetic variability and population structure in the gorgonian coral *Pseudopterogorgia elisabethae* across the Bahamas. *Molecular Ecology* **13**, 2211-2221.
Heist EJ, Gold JR (2000) DNA microsatellite loci and genetic structure of red snapper in the Gulf of Mexico. *Transactions of the American Fisheries Society* **129**, 469-475.
Lee TN, Clarke ME, Williams E, Szmant AF, Berger T (1994) Evolution of the Tortugas gyre and its influence on recruitment in the Florida Keys. *Bulletin of Marine Science* **54**, 621-646.
Lindeman KC, Lee TN, Wilson WD, Claro R, Ault JS (2001) Transport of larvae originating in southwest Cuba and the Dry Tortugas: evidence for partial retention in grunts and snappers. *Proceedings of the 52nd Gulf and Caribbean Fisheries Institute* **52**, 732-747.
Paetkau D, Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology* **4**, 519-520.
Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347-354.
Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.
Piry SA, A., Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536-539.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
Raymond M, Rousset F (1995a) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.
Raymond M, Rousset F (1995b) An exact test for population differentiation. *Evolution* **49**, 1280-1283.
Rosen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365-386. Humana Press, Totowa, NJ.
Schneider S, Kueffer J, Roessli D, Excoffier L (2000) ARLEQUIN: a software for population genetic analysis Version 2.000, Genetics and Biometry Laboratory, University of Geneva, Switzerland.
Shulman MJ, Bermingham E (1995) Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* **49**, 897-910.
Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457-462.
Thorrold SR, Latkoczy C, Swart PK, Jones CM (2001) Natal homing in a marine fish metapopulation. *Science* **291**, 297-299.
Waser P, Strobeck C (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution* **13**, 43-44.
Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**, 388-396.
Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. *Genomics* **7**, 524-530.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
Wright S (1921) Systems of mating. *Genetics* **6**, 111-178.