Preliminary Genetic Stock Assessment of Scamp (Mycteroperca phenax) in Florida Waters

SEDAR68-SID-03

26 July 2019
Updated: 20 September 2019

This information is distributed solely for the purpose of pre-dissemination peer review. It does not represent and should not be construed to represent any agency determination or policy.

Please cite this document as:

Wallace, Elizabeth. 2019. Preliminary Genetic Stock Assessment of Scamp (Mycteroperca phenax) in Florida Waters. SEDAR68-SID-03. SEDAR, North Charleston, SC. 19 pp.

SEDAR 68 South Atlantic and Gulf of Mexico Scamp Data and Assessment Workshop

DRAFT

Working Paper: Preliminary Genetic Stock Assessment of Scamp (*Mycteroperca phenax*) in Florida Waters

Report Prepared By:
Dr. Elizabeth Wallace
Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute
August 29, 2019

Introduction

Scamp are one of 15 grouper species in the genus *Mycteroperca*; their range includes the Gulf of Mexico, Caribbean Sea, and western Atlantic Ocean. Preferred habitat for Scamp (*M. phenax*) consists of reef ledges and rocky bottom at depths up to approximately 100m. They bear close similarity to congener Yellowmouth grouper (*M. interstitialis*). Scamp are protogynous with females reaching maturity at two-five years and the transition to male appearing to be socially controlled (Harris et al. 2002). Spawning occurs from April through August in aggregations of approximately 100 individuals (Manooch et al 1998). Their protogynous life history and the possibility of lek formation in Scamp increase the susceptibility of males to overharvest.

Despite the importance of Scamp as a target species of both recreational and commercial harvest, there is limited biological life history and demographic data available. Similarly, the genetic stock structure of scamp within Florida waters has not been fully evaluated. Previous work by Zatcoff et al. (2004) with a limited six-locus microsatellite panel did not detect population partitioning. This species is currently listed by the International Union for the Conservation of Nature (IUCN) Red List as data deficient. As a result, key information for management of scamp is currently lacking.

The goal of SouthEast Data, Assessment, and Review (SEDAR) 68 is to conduct a stock assessment for Scamp in the Gulf of Mexico and Atlantic Ocean. Through the stock ID webinar series, all available data will be reviewed to determine Scamp stock numbers and locations ahead of the data workshop in March 2020. The purpose of this paper is to present the preliminary results of our genetic stock structure assessment of Scamp, currently underway, so the information may be incorporated into the official working paper for the SEDAR 68 genetics working group.

Methods

Tissue Collection

Scamp specimens were collected from June 2013- January 2019 using a collaborative and opportunistic approach for tissue collection. Fin clips were obtained from other FWC departments whenever possible through independent monitoring, directed research, or fisheries dependent monitoring (angler intercept) activities. Most specimens were collected from along the Florida Gulf coast. When possible fork length (FL, in cm), total length (TL, in cm) and weight (wt, in g) were recorded. FWC staff preserved individual Scamp fin clips in 2mL scintillation vials of 95% ethanol and transferred them to the FWRI genetics department for processing and analysis. As of this reporting, assays have been completed for a total of 556 Scamp specimens collected from five regions, from the Panhandle to the Keys, Florida. The study locations with their respective sample sizes are listed in Figure 1.

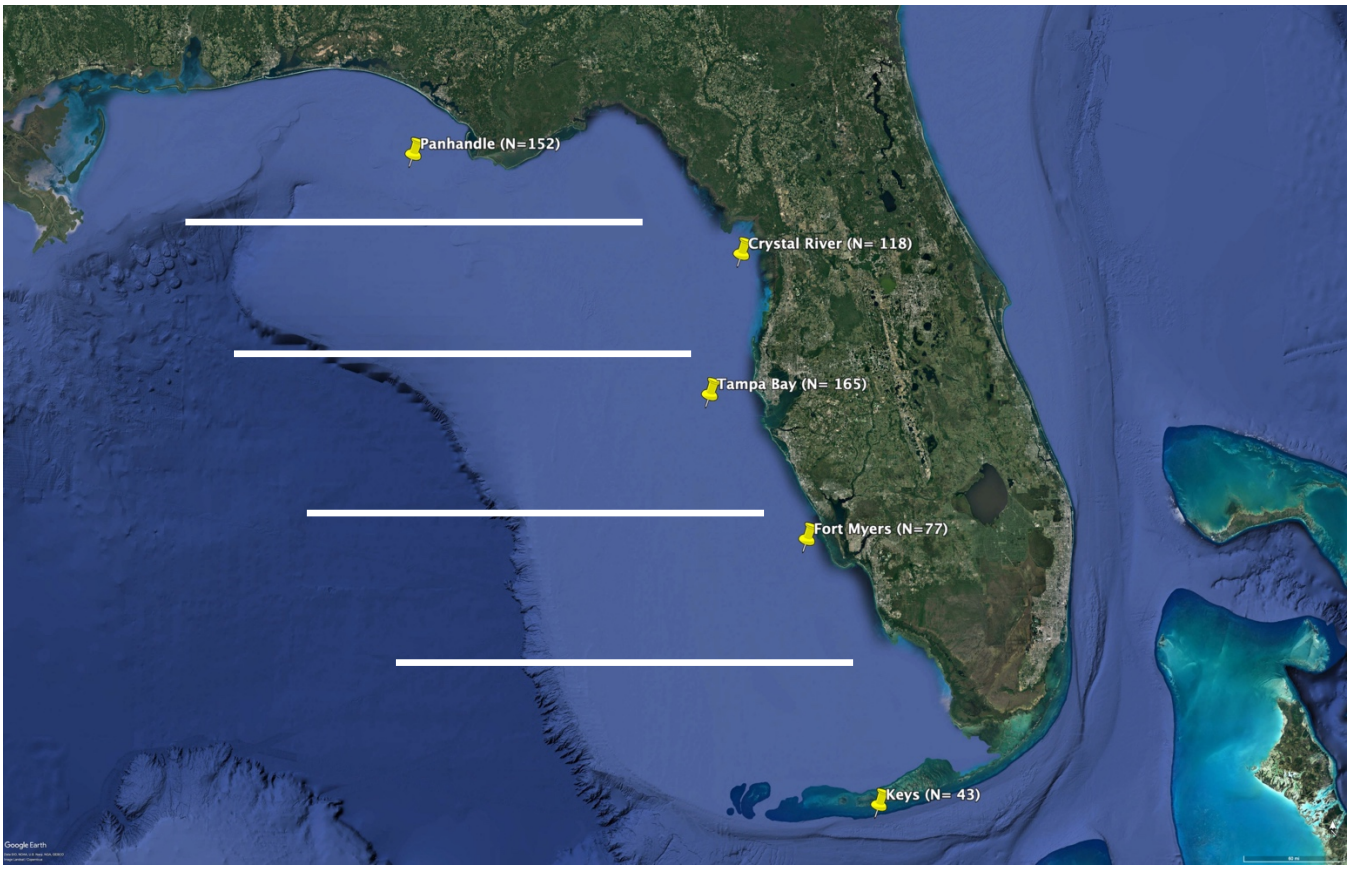


Figure 1. Scamp specimens for the current study were collected from five regions along the eastern Gulf of Mexico, Florida Straits, and Florida Atlantic coast. Sample sizes for each location are listed in parentheses and the white bars approximate regional boundaries.

Lab Procedures

Genomic DNA was purified from tissues through standardized laboratory procedures. The purification protocol utilized the Puregene tissue extraction kit (Gentra Systems, Minneapolis, Minnesota) and followed the manufacturer's directions. Genomic DNA was hydrated to a 100µl final volume in a TE buffer solution. All available Scamp specimens were surveyed with the 31-locus microsatellite array consisting of a combination of published and in-house developed markers (Seyoum et al. 2013, Zatcoff et al. 2004)(Table 1).

Table 1. Microsatellite loci used in the current study.

Locus	Primer Sequence (5′ – 3′) Forward/Reverse	Repeat motif	Allele size range
Emor09	ATAAAGCGTGTGGGTGTACAGTAT ATAATGAACTGCATTCCTTGGGTA	(TG)7/(TG)5	234-238
Emor13	ACACTCGCATGCATACACAAAC AGCACCACTAGGGGACAGTAGTAA	(CA)18	96-118
Emor14-34	CTCCCAAAAGGATTATACAGAGGA AAATGAGTAAGAGGCTTTGATGCT	(AC)6/(CA)22	125-169
Emor20	GACAAGCAGACACACAATTACACA GCAGCACGGTGAATAGAAGTAGTA	(RACA)8	146-162
Emor22	GCATGCTTTTACATGGTTGTGT AGTGGAGATGTGTGTGTGTGTG	(AC)8	104-126
Emor23	GATGTGGCTTTCCTTCTAAAAACA TTTATCTGTAGGCCTATCGTTTGC	(TG)11	139-147
Emor26	CCATAATAAAACCTCCGTTCATGT TCCTTTGTTCCCTTTCAACTTAAC	(GT)19	129-167
Emor28	TCCTGCGTCTTATTTCAAACATTA GAGTGTGAGTGGTGAAAAGATTTG	(TG)14	109-129
Emor31	AAAGACACAGACCAATCCATTTCT GTCTCTAACTGAGATGGATGCAGA	(AC)4/(CA)7	122-124
Emor32	CAGATGAGACCATTTAGAAGCAAA AAGACAACTGATGTGCATTAGCAT	(GT)14	160-164
Emor33	ATTGTCGTTGGGTAGTGTTACAGA GGACATGATCAGTGATATTTGGTG	(GT)11	250-308
Eita03	ATCAGATTTGGAATTAGATCCATTG GTAGGCCAACAGTCTTCTCTATCTG	(TG)25(AG)5	292–322
Eita06	GAATGTACCTTTGCTACAGGAGTGT ATACATTTTTGGTCATGTCTGTGTG	(TR)23	226–256
Eita07	TGCTTGTATATGTGTGTGAGAAAGA TAAGGTCATATTGATCTCCCTGTGT	(TG)16	199–225
Eita12	ATTGTGTAATGTGTTCCTGTTTGTG AGATGGTCTTACGTGTGATCTTGTT	(AGC)9	268–284
Eita13	ACAGAAAGAAGAAAGGTGGGTAAGT CAGAGTCATAAGGGGCAGACTACTA	(TG)10	189–231
Eita19	AAATAACGGGAGTAGCTACCAAGAC CCCCTCACCCGTTTATATTTTATTA	(CA)14	174–218
Eita27	AAGCAGTTCCGATTAGATAACGTC GCTGTGCAGTAAATATGCTTTGAC	(TG)15	172–178
Eita29	ACCTGTATTTCTTAATGCTGATGTCTT GGGAAGTATTCCTTTAACATACCGTAA	(TG)13	156–176
Eita33	TATAGACCAATCAGGTTGCATGAAT ATGAGGGATCATGAAGCAAAATC	(TG)10	156–182
Eita34	TAAATCAGCACATTTATGTAACAAGGT GTTCTCTTTCTCCATGTAGGTGAGTTA	(GA)22	127–157
Eita36	GAGGAAGAAAGACGGAAGATAGACT TTCATGGTTAGTTGATCACTGGAAT	(TG)15	143–159
Eita39	AGCATGTCTAAAGCTGAAGGTGTAT TTTTACTTTGATAGCCAATCTGGAC	(AC)11	100–106
Mbo29	GAGCACGCACACTGAGCAAC TGCCAGTAAGGCAAAGTGGTC	(GT)14	139-201
Mbo48	CAACGTTGTCATAATCTGAGCAT CGTGGATGATGTTAACTTTGGTG	(GACA)11	99-123
Mbo66	CGCATGTTTGTAAGAACAGGAAG GCTTCACTCTTGGGTTGTTGG	(GT)8	102-112
Gag45	CCTCACGACGAGTCCAGGAG GTTTGCCTTAACGGATGTCTTTCT	(GT)14	113-131

Screening for the microsatellite array was conducted in multiplexed polymerase chain reactions (PCRs) with 3 to 5 pairs of fluorescently labeled primers in 12.5µl final volume assays and approximately 100ng of genomic DNA. These reactions contained 0.07µl GoTaq polymerase, 0.01-0.04µl combined forward and reverse 100µM primers, 0.3µM each dNTPs, 2.32mM

MgCl2, 2.33µl 5x GoTaq buffer, 0.06µl BSA, and 7.42µl sterile distilled water. Target loci were amplified under the following touchdown thermal cycling conditions: 94°C for 0:45; 5 cycles of 94°C for 0:45, 58°C for 0:45, 72°C for 0:45; 8 cycles of 94°C for 0:40, 57°C for 0:40, 72°C for 0:40; 10 cycles of 94°C for 0:35, 56°C for 0:35, 72°C for 0:35; 15 cycles of 94°C for 0:30, 55°C for 0:30, 72°C for 0:30; and a final 72°C extension for 5:00 followed by a 4°C hold.

Following amplification, an aliquot of 1.5µl each PCR product was mixed with 11µl HiDi formamide and 0.18µl of a custom ROX labeled size standard prepared in-house following the DeWoody et al. (2004) protocol. These products were briefly denatured at 95°C for 2:30 minutes and flash cooled on ice. Microsatellite fragment screening was conducted on an Applied Biosystems 3130XL genetic analyzer and genotypes scored with Genemapper software.

Due to the possibility of misidentification and uncertain species boundaries between Scamp and Yellowmouth, available specimens (N=9) collected from the Florida West coast and field identified as Yellowmouth by FWC staff were screened with the Scamp microsatellite panel for comparative analysis. Additionally, these specimens were sequenced for a 622bp segment of the mtDNA COI gene using the FishF2 and FishR2 primers of Ivanova et al. (2007) following standard methods described in Wallace and Tringali (2016).

Genetic Species Identification Analyses

The combined Scamp and presumptive Yellowmouth microsatellite dataset (N=565) was rigorously evaluated with spatial correlation and Bayesian clustering methodologies. The dataset was first visualized in the three-dimensional spatial correlation analysis program *Genetix* (Belkhir et al. 2000) to identify outliers. Then Bayesian categorical tests were run in *Newhybrids* (Anderson and Thompson 2002) to evaluate species or hybrid assignments. Additionally, the nine presumptive Yellowmouth COI sequences were edited and aligned in Geneius© (Biomatters LTD.) and compared against available sequences in the National Center for Biotechnology Information Genbank through a nucleotide BLAST search.

Genetic Diversity and Clustering Analyses

To assess patterns of genetic diversity and differentiation, several genetic indices, a genotype curve, and a locus specific allelic table were calculated in the r statistical framework using the *poppr* and *adegenet* packages (Kamvar et al. 2019, Jombart et al. 2018). Microsatellite loci were evaluated comparatively for conformance to Hardy Weinberg Proportions (HWP) with package *diveRsity* (Keenan et al. 2013) and in the Excel based program GenAlex (Peakall et al. 2006). Linkage disequilibrium (LD) was assessed through index of association with 1000 permutations in packages *poppr* and *magrittr* (Bache and Wickham 2014). Interpretation of HWP and LD analysis results followed Waples (2015).

The presence of private alleles within study locations was assessed with the r package *poppr*. For these evaluations of the data, Scamp specimens were partitioned into five ad hoc groups according to collection region. Genetic differentiation was further explored among assigned collection locations with Hedrick's G'st, Jost's D, and Fst analyses in the r package *diveRsity* and in GenAlex. These estimators range from zero (absence of genetic variation) to one (complete differentiation). Though both G'st and Jost's D evaluate allele frequency differences among populations, they measure slightly different aspects of allelic variation. The G'st estimator measures genetic distance between populations and is standardized by dividing by the maximum possible Gst according to the heterozygosity of the locus. The genetic differentiation estimator Jost's D measures the allelic variation fraction among populations. The *diveRsity*

calculated metrics report bias corrected confidence intervals to account for known potential upward bias (but do not report Bonferroni corrected p-values), while the GenAlex metrics report classical Bonferroni corrected p-values. These metrics were evaluated comparatively.

Spatial separation of population clusters was assessed through a discriminant analysis of principal components (DAPC) test in *adegenet*. This multivariate analysis creates synthetic variables in an ANOVA framework through linearized Fst values and provides a three-dimensional scatter plot output of all individuals. We selected the number of discriminant functions based on lowest Bayesian information criterion (BIC) score and implemented the alpha optimization method to determine the number of principal components retained, ensuring that observed clustering patterns were derived analytically and not driven by ad hoc cluster assignments.

Population partitioning in Scamp was further assessed using the Bayesian statistical framework implemented in the program *Structure* (Pritchard et al. 2000). Variation in the number of identified population partitions and assignment of scamp specimens to those partitions was evaluated in two sets of *Structure* analyses. The modeling options implemented were varied across the analysis sets to identify data sensitivities to particular parameters in the model framework:

Set 1: Admixture model with the locprior and correlated allele frequencies options implemented.
Set 2: Admixture model with only the allele frequencies correlated option implemented.

The *Structure* analyses were run on the Brigham Young University Fulton Supercomputing cluster under the following Marcov chain Monte Carlo (MCMC) conditions: 100,000 step burn-in period followed by 5,000,000 steps of data collection, the migration prior set to 0.01, and 1-6 maximum populations (K). The sets were each replicated 7 times and the results compiled for population pattern inference. The compiled results were then evaluated for ad hoc inference of population partitioning using both the L(K) and ΔK methods in *Structure Harvester* (Earl & VonHoldt 2012). Due to the potential for group label switching among replicates, the program *Clumpp* was used to compile the group assignments for the inferred number of populations (Jakobsson et al. 2007), and final bar plots created in *distruct* (Rosenberg 2004). In addition to the total dataset, a subset (N=470) containing non-spawning individuals collected July- December across sampling years was assessed through pairwise comparisons, DAPC, and Bayesian clustering methods to identify any potential seasonal influence. Upon identification of discrete clusters, specimen removal was employed on the total and non-spawning dataset in an effort to obtain K=1 and five replicates were completed under the set one Structure model conditions.

Results

Specimen collection data was evaluated for potential length or weight bias, as well as the potential for any collections that may have inadvertently sampled family groups or a single cohort. Length and weight data were not available for all specimens, however recorded lengths ranged from 178-770mm FL (mean 644mm, N=433) and weights ranged from 0.575-8.2lb (mean 1.75lb, N=97). No obvious indications of bias were found, and the nature of collections (infrequent, low volume) make family group/ cohort sampling within any region unlikely.

Genetic Species Identification Analyses

Spatial correlation analysis based on the 28-locus microsatellite panel identified 6 outliers consisting of two presumptive Yellowmouth and four presumptive Scamp (Figure 2).

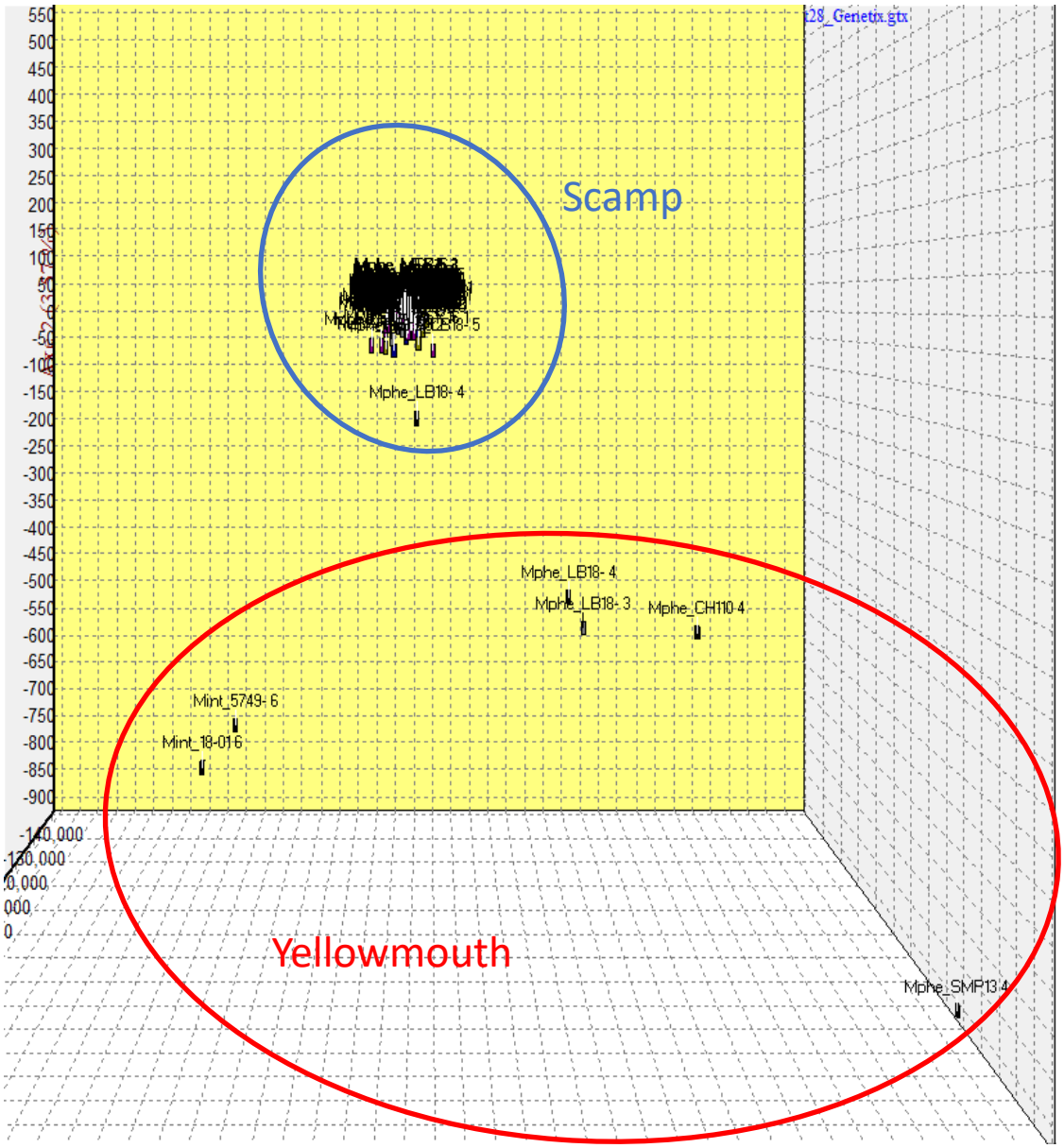


Figure 2. Spatial clustering *Genetix* results identified six outlier specimens.

Subsequent probabilistic *NewHybrids* analyses assigned nine specimens as Yellowmouth. These specimens consisted of three presumptive Yellowmouth (two of which were also *Genetix* outliers) and six presumptive Scamp, providing evidence for misidentifications. Based on the total dataset, the observed misidentification rate was 1.1%. Four specimens were inconclusive, receiving low probability assignments for most categories. The rest of the specimens (presumptive Scamp) were assigned to the second-generation hybrid (F2) category, rather than

pure Scamp. This can occur when population structure exists within a species, as *NewHybrids* operates in a similar manner to that of *Structure* (i.e. sorting occurs based on allele frequency patterns), but with the model assumption that no structure exists within either species. In contrast, *Structure* assumes a single species with variability solely due to population structure within it. Alternatively, the results may reflect the absence of distinct species in the dataset, resulting in the inability to sort and assign Scamp individuals to the expected pure species category.

To further evaluate the F2 assignments, the data was partitioned according to *Structure* assigned clusters (detailed below, based on set 1 model conditions) and run in *NewHybrids* separately. Presumptive Scamp individuals in the smallest population cluster (N=8) were assigned as pure Scamp with high probabilities (>0.99), with all nine presumptive Yellowmouth assigned as such. Results were less clear and more stochastic for the third *Structure* cluster. Forty-two individuals were assigned as Yellowmouth with significant probabilities (>0.95). This set included two mis-identified in the field as Scamp and assigned in *Structure* to population cluster three along with other Scamp specimens. Another 50 individuals were assigned as Scamp and 19 as F2s with significant probabilities. The rest of the specimens in this partition were assigned with non-significant probabilities as Yellowmouth, Scamp, or F2 indicating difficulty assigning specimens as expected under a two species scenario.

Results from the COI sequencing identified one of the nine presumptive Yellowmouth as such, according to a Genbank BLAST search. Two specimens did not sequence well, however microsatellite-based analyses identified them as Yellowmouth. The remaining four individuals matched a Genbank submission for Broomtail grouper (*M. xenarcha*) found along the US Pacific coast, while the second closest match was for Yellowmouth as no Scamp COI sequence data is available in the database.

Genetic Diversity and Clustering Analyses

In general, the microsatellite data reflected moderately high genetic diversity (Table 2). Allelic richness ranged from three to 22 (mean = 9.571) among the 28 loci and the observed per locus heterozygosity (Ho) ranged from 0.02- 0.91 (mean = 0.5). Among the five collection regions, Nei's unbiased gene diversity varied little, ranging from 0.512 for Crystal River to 0.553 for the panhandle. Four of the five collection regions had a higher than expected number of multilocus genotypes (MLG), after correction for sample size variance through rarefaction (Table 2).

Table 2. Locus (A) and collection location (B) specific allelic diversity parameters were calculated in *poppr* and *adegenet*. Observed heterozygote deficiencies are in bold.

A)

locus	allele	Simpsons Index	Ho	Hexp	Evenness
Emor09	4	0.247	**0.21**	0.25	0.548
Emor13	4	0.307	0.31	0.31	0.624
Emor14-34	11	0.572	**0.52**	0.57	0.508
Eita39	8	0.584	**0.54**	0.59	0.631
Emor20	3	0.418	**0.23**	0.42	0.845
Emor28	10	0.705	**0.69**	0.71	0.779
Eita03	19	0.898	0.91	0.90	0.853
Emor22	7	0.069	**0.05**	0.07	0.347
Emor23	10	0.552	**0.52**	0.55	0.663
Emor26	12	0.798	**0.76**	0.80	0.76
Emor31	3	0.036	**0.02**	0.04	0.349
Emor32	4	0.05	**0.04**	0.05	0.358
Emor33	16	0.864	**0.82**	0.87	0.783
Mbo48	9	0.717	**0.42**	0.72	0.713
Emor38	4	0.183	**0.03**	0.18	0.447
Eita12	3	0.501	0.50	0.50	0.986
Mbo29	19	0.866	**0.85**	0.87	0.733
Mbo66	6	0.511	**0.50**	0.51	0.815
Eita06	6	0.398	**0.37**	0.40	0.633
Eita34	6	0.63	**0.62**	0.63	0.799
Eita36	9	0.599	**0.58**	0.60	0.587
Eita27	22	0.92	**0.91**	0.92	0.853
Eita13	22	0.915	**0.88**	0.92	0.814
Gag45	5	0.24	**0.22**	0.24	0.545
Eita33	5	0.218	**0.21**	0.22	0.451
Eita19	6	0.544	**0.54**	0.55	0.778
Eita7	17	0.832	0.83	0.83	0.703
Eita29	18	0.857	0.86	0.86	0.738
mean	**9.571**	**0.537**	**0.50**	**0.54**	**0.666**

B)

Pop	N	MLG	eMLG	SE	Gene Diversity
Panhandle	153	153	43	1.61E-05	0.553
Crystal River	118	118	43	5.93E-06	0.512
Tampa Bay	165	165	43	0.00E+00	0.528
Fort Myers	77	76	42.7	4.62E-01	0.546
Keys	43	43	43	0.00E+00	0.541
Total	**556**	**555**	**43**	**7.63E-02**	**0.538**

Many loci (17 of 28) were identified as having significant departures from expected Hardy Weinberg proportions (HWP), according to an exact test. The 11 adhering to expected HWP across all collection regions were Emor13, Eita03, Eita12, Eita13, Mbo29, Eita 27, Eita34, Eita36, Eita19, Eita7, and Eita29. Closer examination of the data revealed slight to substantial heterozygote deficiencies in most loci. The observed heterozygosities were equal to or higher than expected (He) in: Emor13, Eita03, Eita12, Eita7, and Eita29 (these loci also met HWP expectations). The observed heterozygote deficits may be due to null alleles, the Wahlund effect, age structure, or assortative mating. Evaluation of Fis/ Fst indicated a positive correlation, which supports population structure as the causative force (Figure 3).

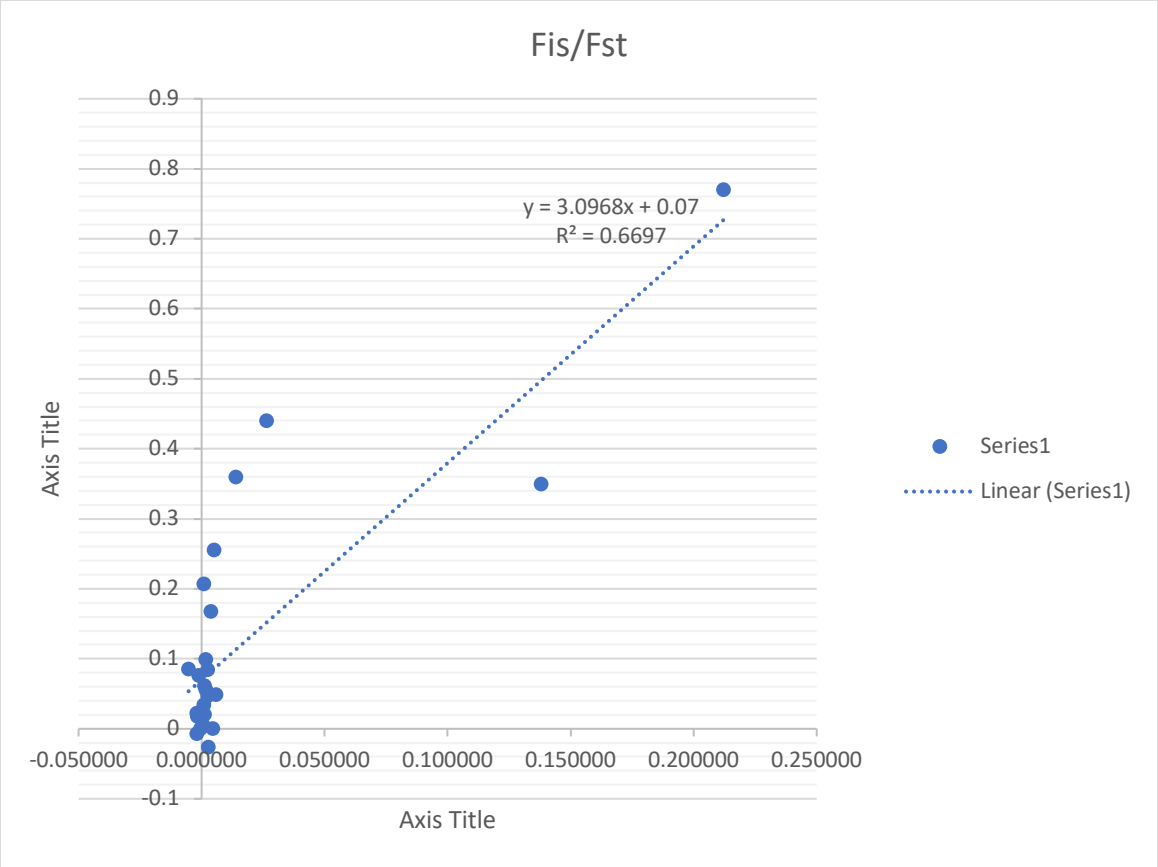


Figure 3. Evaluation of the inbreeding coefficient Fis to the fixation index (Fst) identified a positive correlation, supporting Wahlund effect as the source of HW deviations.

Patterns of HW deviation among collection regions was further evaluated in *diveRsity* and *GenAlex*. The results were similar, with only a few identified as significant departures in one analysis but not the other (Table 3). Further, only two loci (Emor20 and Mbo48) displayed departures across all regions. Due to the numerous HW departures, the K=1 cluster identified in Structure (described below) was also examined separately. Though fewer departures were observed, and none occurred across all regions, many (N=44) remained (data not shown). The global linkage disequilibrium test of independent association (rd=0.104, P=0.977) indicated no linkage among loci. The observed patterns are not surprising given the available microsatellite loci were developed for other groupers. Further reduction of the dataset to include only conforming loci would result in a loss of power to resolve biologically informative population patterns.

Table 3. Evaluation across collection regions and loci for HW conformance using *diversity* (A) and *GenAlex* (B). Significant (p-values) is for each method is highlighted in yellow, while discrepancies between methods are shown in orange. Location code: 1= Panhandle, 2= Crystal River, 3= Tampa Bay, 4= Fort Myers, 5= Keys.

A)

	1	2	3	4	5
Emor09	0.0002489	0.2837489	0.0000000	0.0000000	0.8449490
Emor13	0.8536362	0.1401565	0.8662256	0.2607324	0.6951018
Emor14-34	0.0347130	0.9231828	0.3412683	0.0000000	0.8255265
Eita39	0.4930939	0.8264364	0.0000004	0.8913888	0.8348763
Emor20	0.0000086	0.0090445	0.0000565	0.0000392	0.0000057
Emor28	0.0000000	0.0756602	0.1072911	0.0442523	0.8903259
Eita03	0.2423308	0.9459028	0.1210703	0.9985929	0.9674592
Emor22	0.0000000	1.0000000	0.0000000	0.0000000	0.8274351
Emor23	0.0000000	0.6308984	0.0000000	0.0000000	0.8466782
Emor26	0.0002551	0.0045805	0.0127044	0.9546312	0.2495527
Emor31	0.0000000	1.0000000	0.8569085	0.0012649	0.0036030
Emor32	0.7778501	0.9523468	0.0000000	0.0000008	0.8897821
Emor33	0.9549127	0.4250250	0.4153224	0.0000001	0.9625300
Mbo48	0.0000000	0.0000000	0.0000000	0.0000561	0.0034695
Emor38	0.0000000	0.9574934	0.0000000	0.0000000	0.9272742
Eita12	0.5827116	0.5884079	0.6796429	0.4752403	0.2606791
Mbo29	0.0127952	0.4531529	0.4163578	0.5625497	0.6706690
Mbo66	0.0000000	0.8727146	0.0000000	0.9066162	0.4770505
Eita06	0.0000000	0.8789145	0.4951827	0.0000007	0.0040065
Eita34	0.7810569	0.4642728	0.9890473	0.3957186	0.0223389
Eita36	0.8400958	0.0698711	0.8545366	0.1678093	0.0055601
Eita27	0.0000000	0.9650717	0.4876411	0.2372198	0.9699533
Eita13	0.1738515	0.0000108	0.0000135	0.9632734	0.1091480

Gag45	0.0000000	0.5341445	0.0189982	0.0000000	0.9723215
Eita33	0.0000000	0.9292027	0.9829842	0.9540263	0.0080986
Eita19	0.1587921	0.0126285	0.5321233	0.6113972	0.9148790
Eita7	0.6075792	0.0037586	0.4437941	0.4654706	0.9534962
Eita29	0.8629225	0.8940858	0.0000001	0.0302369	0.3369077

B)

	1	**2**	**3**	**4**	**5**
Emor09	0.007	0.884	0.000	0.000	0.820
Emor13	0.037	0.531	0.739	0.000	0.695
Emor14-34	0.000	0.569	0.030	0.000	0.986
Eita39	0.684	0.140	0.000	0.703	0.254
Emor20	0.000	0.000	0.000	0.000	0.000
Emor28	0.000	0.209	0.001	0.005	0.990
Eita03	0.000	0.399	0.219	0.999	0.966
Emor22	0.000	monomorphic	0.000	0.000	0.827
Emor23	0.000	0.593	0.000	0.000	0.311
Emor26	0.018	0.004	0.163	0.813	0.190
Emor31	0.000	0.272	0.955	0.000	0.003
Emor32	0.000	1.000	0.000	0.000	0.977
Emor33	0.999	0.795	0.339	0.037	0.897
Mbo48	0.000	0.000	0.000	0.000	0.000
Emor38	0.000	1.000	0.000	0.000	0.927
Eita12	0.567	0.865	0.000	0.000	0.585
Mbo29	0.146	0.776	0.969	0.011	0.862
Mbo66	0.000	0.472	0.000	0.000	0.998
Eita06	0.000	0.890	0.495	0.000	0.004
Eita34	1.000	0.840	0.706	0.553	0.135
Eita36	0.261	0.246	0.952	0.422	0.019
Eita27	0.000	0.776	0.970	0.883	0.503
Eita13	0.023	0.004	0.000	0.696	0.234
Gag45	0.000	0.614	0.009	0.000	0.914
Eita33	0.000	0.015	0.983	0.999	0.008
Eita19	0.363	0.312	0.539	0.808	0.238
Eita7	0.797	0.002	0.545	0.593	0.678
Eita29	0.156	0.909	0.075	0.149	0.177

Relatively few private alleles were identified in the Scamp microsatellite data (Figure 4). After correction for sample size variation through rarefaction, most private alleles were identified from the panhandle, Tampa Bay, and Fort Myers collection regions.

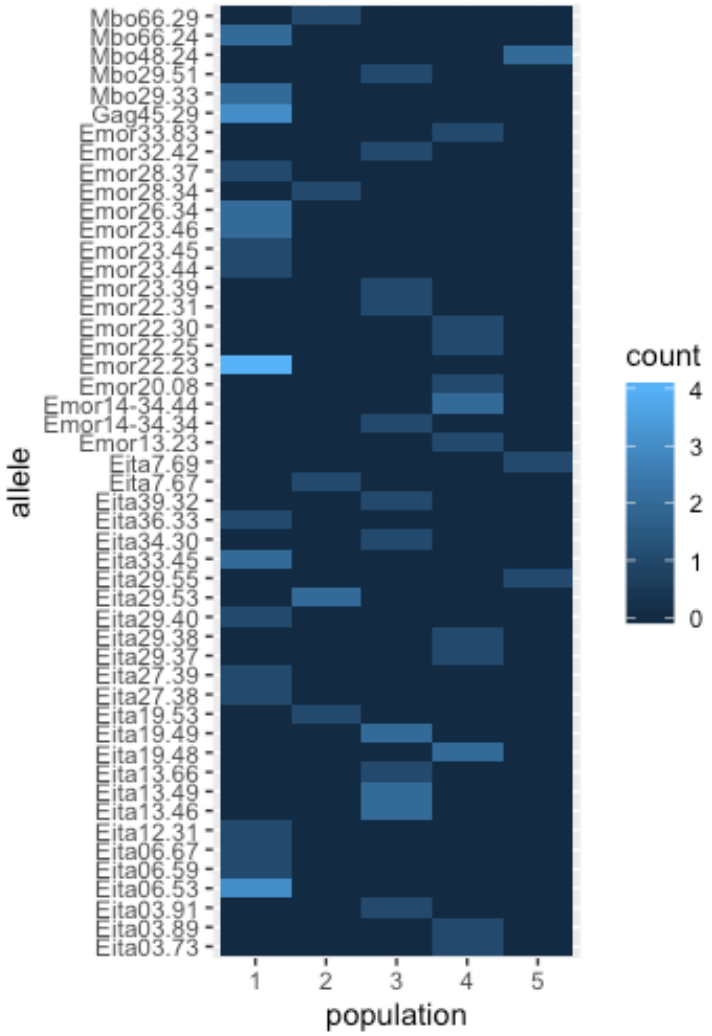


Figure 4. Heat map displaying the presence of private alleles, corrected for sample size variation through rarefaction, within collection regions.

Pairwise genetic metrics (G'st, Josts D, and Fst) between collection regions were analyzed in *diveRsity* and *GenAlex*, and results evaluated comparatively. Results from *diveRsity* identified significant differentiation (G'st > 0.05) (Table 3A) in all panhandle comparisons as well as between Crystal River and Tampa Bay. Fst values displayed the same pattern (results not shown), while Josts D did not indicate significant differentiation (Table 3B). Metrics calculated in *GenAlex* revealed a similar pattern, though with the Josts D and Fst estimators rather than G'st. Following Bonferroni correction, Josts D and Fst p-values (all 0.001) indicated significant differentiation for panhandle comparisons and Crystal River to Tampa Bay. The observed pattern of genetic differentiation between all panhandle comparisons and that of Crystal River and Tampa Bay also held for the reduced non-spawning dataset.

Table 3. Genetic differentiation among assigned collection locations was assessed with A) G'st and B) Jost's D analyses in R package *diveRsity*. Lower triangle values are the estimator, and upper triangle values are the lower 95% confidence interval bound. Significant comparisons highlighted in yellow. Location code: 1= Panhandle, 2= Crystal River, 3= Tampa Bay, 4= Fort Myers, 5= Keys.

A)

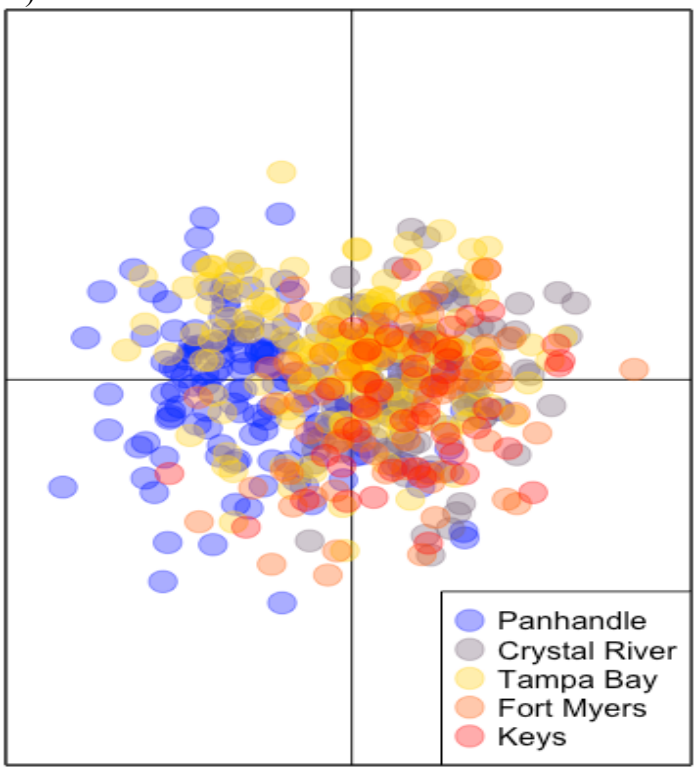
	1	**2**	**3**	**4**	**5**
1	~	0.029	0.0065	0.0145	0.0135
2	0.05762283	~	0.0016	-0.0091	-0.0164
3	0.02423777	0.01451284	~	-0.0031	-0.0091
4	0.04063048	0.00486562	0.0091261	~	-0.0229
5	0.04307613	0.00377085	0.01198385	-0.006599	~

B)

	1	**2**	**3**	**4**	**5**
1	~	-0.0003	-0.0024	-0.0021	-0.0043
2	0.0024	~	-0.0021	-0.0047	-0.0061
3	0.0011	0.0005	~	-0.0038	-0.0059
4	0.0017	0.0003	0.0002	~	-0.006
5	0.0007	0	0.0001	-0.0002	~

The DAPC multivariate analysis was run for four discriminant functions and 25 principal components, as indicated through lowest BIC value and alpha optimization score, respectively. These results display one distinct cluster containing Scamp from all five regions (Figure 5A). Though a few outliers are present, none of them suggest the presence of a distinct, separate population. Yet within the cluster, there is some indication of subtle North- South separation along the Gulf of Mexico Florida coast. Most of the Scamp specimens collected from the panhandle (Alabama to Apalachee Bay) are confined to the left quadrants, while most Scamp from southern locations are contained in the right quadrants. The non-spawning partition yielded a similar result (based on four discriminant functions and 22 principal components) (Figure 5B). The observed pattern suggests subtle isolation by distance for Scamp across the study area examined.

A)



B)

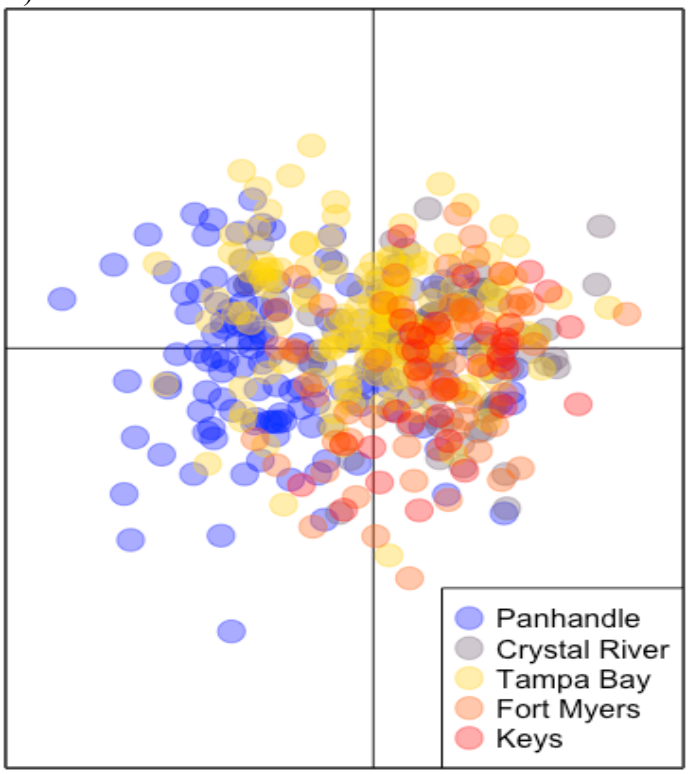
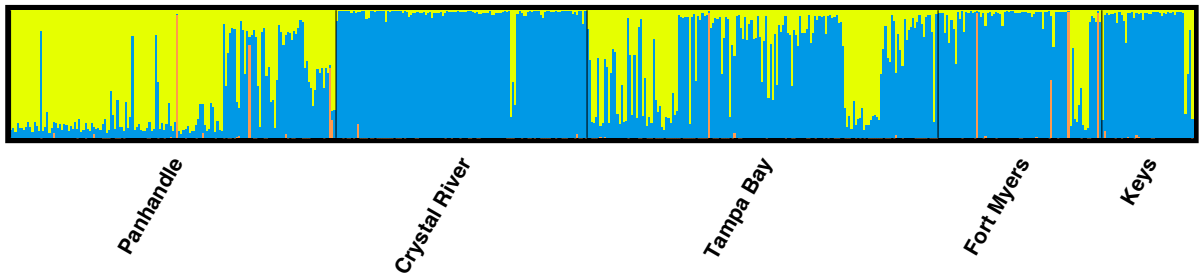


Figure 5. Discriminant analysis of principal components (DAPC) results, assessed in *adegenet,* for the total dataset (A) and non-spawning dataset (B).

The results from the population clustering analyses on the total dataset in *Structure* were consistent between set one and two modeling frameworks and between the L(K) and ΔK inference methods, suggesting three genetic clusters (Figure 6). However, the third cluster consisted of only eight individuals under set one conditions (collected from the panhandle, Tampa Bay, and Fort Myers). The same eight specimens were assigned to the third cluster under set two conditions. Otherwise, specimen assignments under set two parameters resulted in more individuals assigned to mixed ancestries with low probabilities (<0.95) when locprior was not implemented. The six suspected Yellowmouth mis-identified as Scamp were assigned to clusters one (N=4) and two (N=2).

A)



B)

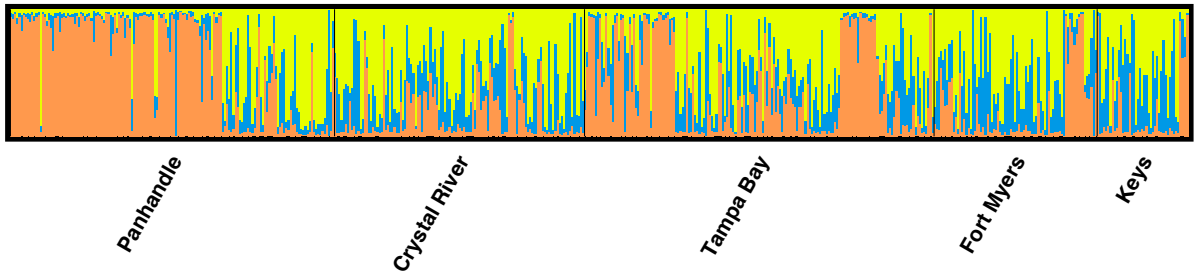
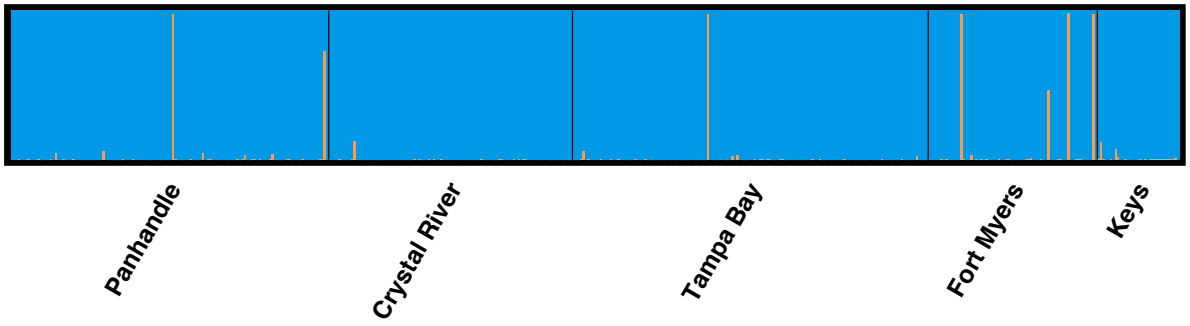


Figure 6. *Structure* bar plots for K=3 under set one (A) and set two (B) modeling framework on the total dataset, displaying individual assignments as vertical lines composed of proportional ancestry.

Results from the population clustering analyses on the non-spawning data partition were not consistent between set one and two modeling frameworks; however, were consistent between the L(K) and ΔK inference methods within each set. With locprior implemented, *Structure* analyses suggesting two genetic clusters, while set two conditions supported three (Figure 7). The few specimens assigned to cluster two in set one included five of the six suspected Yellowmouth mis-identified as Scamp.

A)



B)

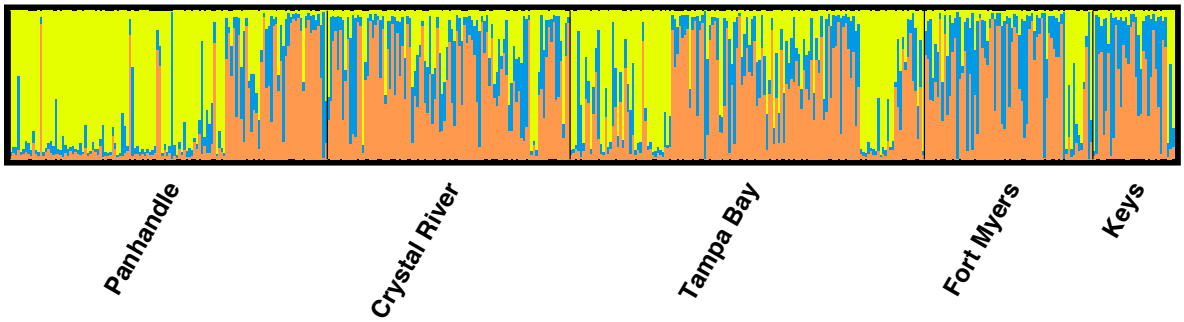


Figure 7. *Structure* bar plots for K=2 under set one (A) and K=3 for set two (B) modeling framework on the non-spawning data partition, displaying individual assignments as vertical lines composed of proportional ancestry.

Under set one conditions, individuals assigned to the second and third clusters were removed and cluster one was re-analyzed alone. Barring further subtle population structure, this should yield K=1. Instead, the inferred number of genetic clusters was 2 for both L(K) and ΔK approaches. Eight specimens were assigned to the second cluster; these included four suspected Yellowmouth mis-identified as Scamp (individuals that were assigned to cluster one in the total dataset).

The variable number of inferred populations in these results suggests subtle population differentiation exists across the study area and/or allele frequency variation due to isolation by distance. Applying the locprior option resulted in stronger assignments and provide a clearer pattern. When considered with the DAPC and pairwise comparisons, results suggest population structure though not strictly geographic in nature. There does appear to be one cluster dominating the panhandle that has decreasing frequency moving South. The exception is the Crystal River region which displays cluster two membership under set one conditions. The models implemented in *Structure* do not handle isolation by distance data well, and it may lead to arbitrary numbers of inferred populations and a majority of specimens being assigned as mixed ancestry.

Species level questions

The *NewHybrids* and *Structure* results highlight the limitations inherent in both methods, and the lack of any microsatellite-based method capable of accounting for both potential sources of allele frequency variation (species and population level) simultaneously. Based on the combined species identification and population results, the validity of Scamp and Yellowmouth as distinct species remains uncertain. The best course of action is to conduct a rigorous coalescent multi-locus species delimitation study to evaluate boundaries among Scamp, Yellowmouth, and Broomtail Groupers. We will initiate this work as soon as additional presumptive Yellowmouth specimens are obtained from the western GOM, as well as Broomtail representatives.

Citations

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*(3), 1217-1229.

Bache, S. M., & Wickham, H. (2014). magrittr: a forward-pipe operator for R. *R package version*, *1*(1).

Belkhir, K., Borsa, P., Goudet, J., Chikhi, L., & Bonhomme, F. (2000). GENETIX, version 4.05. Laboratoire Genome, Populations. *Interactions, CNRS UPR*, *9060*.

DeWoody, J. A., Schupp, J., Kenefic, L., Busch, J., Murfitt, L., & Keim, P. (2004). Universal method for producing ROX-labeled size standards suitable for automated genotyping. *Biotechniques*, *37*(3), 348-352.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), 359-361.

Harris, P. J., Wyanski, D. M., White, D. B., & Moore, J. L. (2002). Age, growth, and reproduction of scamp, Mycteroperca phenax, in the southwestern North Atlantic, 1979–1997. *Bulletin of Marine Science*, *70*(1), 113-132.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., & Hebert, P. D. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, *7*(4), 544-548.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801-1806.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070-3071.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782-788.

Manooch, C. S., Potts, J. C., Burton, M. L., & Harris, P. J. (1998). Population assessment of the scamp, Mycteroperca phenax, from the southeastern United States.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419-420.

Peakall, R. O. D., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular ecology notes*, *6*(1), 288-295.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular ecology notes*, *4*(1), 137-138.

Seyoum, S., Tringali, M. D., & Sullivan, J. G. (2005). Isolation and characterization of 27 polymorphic microsatellite loci for the common snook, Centropomus undecimalis. *Molecular Ecology Notes*, *5*(4), 924-927.

Wallace, E. M., & Tringali, M. D. (2016). Fishery composition and evidence of population structure and hybridization in the Atlantic bonefish species complex (Albula spp.). *Marine biology*, *163*(6), 142.

Waples, R. S. (2014). Testing for Hardy–Weinberg proportions: have we lost the plot?. *Journal of Heredity*, *106*(1), 1-19.

Winter, D. J. (2012). MMOD: an R library for the calculation of population differentiation statistics. *Molecular ecology resources*, *12*(6), 1158-1160.

Zatcoff, M. S., Ball, A. O., & Sedberry, G. R. (2004). Population genetic analysis of red grouper, Epinephelus morio, and scamp, Mycteroperca phenax, from the southeastern US Atlantic and Gulf of Mexico. *Marine Biology*, *144*(4), 769-777.