Stock Structure, connectivity, and effective population size of red snapper (*Lutjanus campechanus*) in the U.S. waters of the Gulf of Mexico

David S. Portnoy

SEDAR74-RD56

March 2021

STOCK STRUCTURE, CONNECTIVITY, AND EFFECTIVE POPULATION SIZE OF RED SNAPPER (*LUTJANUS CAMPECHANUS*) IN U.S. WATERS OF THE GULF OF MEXICO

David S. Portnoy
Marine Genomics Laboratory
Department of Life Sciences
Texas A&M University
Corpus Christi, Texas 78412

FINAL REPORT

MARFIN GRANT PROGRAM GRANT NO NA12NMF4330093

NOVEMBER 2017

I. EXECUTIVE SUMMARY

Three concurrent subprojects were completed with the common goal of providing information about stock structure and genetic demography of Gulf red snapper using a cutting-edge, next-generation sequencing approach.

Subproject 1: Using double digest restriction-site associated DNA (ddRAD) sequencing in non-model, exploited marine fishes: Reference assembly and post-assembly processing were examined to understand how to efficiently produce sets of reliable and repeatable SNPs for non-model marine fishes. In the course of this subproject, a variant calling pipeline specifically for population genomic applications was designed and a manuscript describing the approach has been published. Further, methods for SNP filtering and SNP thinning using haplotyping were explored and a manuscript describing the technique was published. Both publications, which describe the results of this subproject in detail, have been attached as supplementary documents. In addition, two manuscripts that further explore best practices for using RAD sequencing in a population genomic context are in preparation. The results of this subproject are invaluable, providing a resource for future genomics studies of marine fishes, as few assembled and annotated genomes exist for exploited species.

Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. (2014) *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ,* 10: e431.

Willis, S. C., Hollenbeck C. M., Puritz, J. B., Gold, J. R., Portnoy D. S. (2017) Haplotyping RAD loci: an efficient method to filter multi-copy loci and account for linkage disequilibrium. *Molecular Ecology Resources*, 17: 955-965.

Puritz, J. B., Gold, J. R., Portnoy, D. S. (2017) The effects of over-splitting loci on population genetic inference from Rad sequencing. In preparation for *Molecular Ecology Resources.*

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., Gold, J. R., Portnoy D. S. (2017) These are not the loci you are looking for: Principle of effective SNP filtering for

the molecular ecologist. In preparation for *Molecular Ecology*.

Subproject 2: Partitioning of genomic variation in young-of-the-year red snapper: Patterns of genomic variation were assessed within and among young-of-the-year (YOY) red snapper samples collected at six natural and two artificial reef sites in the northern Gulf of Mexico, using 7,382 SNPs. Small but significant levels of genetic heterogeneity were detected between geographic samples, including samples less than 5 kilometers apart. Estimates of the effective number of breeders varied between locations and point estimates ranged from 1,281-59,110. Estimates of relatedness within samples were low. Significant genetic heterogeneity also was detected between two groups of adults, one sample off Louisiana and the other off northern Florida. There was no clear relationship between the genetic similarity of pairs of adult and juvenile samples and the geographic distance between them. Taken as a whole these data do not support the notion that genetic differences between samples of YOY results from sweepstakes recruitment, in which very small numbers of breeders contribute to recruitment; instead they demonstrate that within a single year of recruitment, YOY red snapper originate from genetically distinct groups of spawning adults. This dynamic is consistent with metapopulation-like structure previously hypothesized to be present in Gulf red snapper. The results of this subproject have been published and that manuscript is attached as a supplementary document.

Puritz, J. B., Gold, J. R., and Portnoy, D. S. (2016) Fine-scale partitioning of genomic variation among recruits in an exploited fishery: causes and consequences. *Nature: Scientific Reports*, 6, 36095.

Subproject 3: Populations structure of red snapper in the U.S Atlantic and Gulf of Mexico: Patterns of diversity were assessed within and among 11 geographic samples of mixed-age red snapper. Levels of within sample diversity were similar among samples and 18 outlier loci, putatively under directional selection, were identified. All tests of global heterogeneity were significant but estimates of pairwise F_{ST} for neutral and outlier data sets did not reveal interpretable patterns. Spatial analysis of principal components (sPCA) indicated global structuring and suggested that samples were best grouped into four regions (Carolinas, Florida,

western Gulf and southern Gulf) or two regions (Carolinas and Florida, western and southern Gulf) depending on the connection network used. The four-region model was supported by discriminant analysis of principle components (DAPC) and estimates of pairwise F_{ST} between the four regions were significant for the outlier data set but not for the neutral data. Similarly, for the two-region model, estimates of pairwise F_{ST} were significant for the outlier dataset and not significant for the neutral data set. Point estimates of contemporary effective size ranged from 40,841 for the combined Carolinas and Florida to 65,733 for all geographic samples combined (global). Estimates of migration using two methodologies suggested rates generally below 10% and favored movement into Florida. In sum, the data support the conclusions of previous studies that red snapper are not genetically homogenous throughout U.S. waters. Though the data here do suggest that the Gulf of Mexico may be comprised of two stocks there was not a strong consensus across analyses based on the neutral data set. This may be due to non-equilibrium conditions in Gulf red snapper, i.e. recent range expansion associate with the end of the last glacial period, and/or fairly high levels of connectivity associated with proposed metapopulation structure. Further, genomic analysis of stock structure is warranted but will likely require sampling that takes into account age-structuring of red snapper and ontogenetic differences in habitat use.

II. PURPOSE

Red snapper, *Lutjanus campechanus*, is a long-lived, shelf-spawning species that supports economically important commercial and recreational fisheries in the Gulf and U.S Atlantic (Hood and Strelcheck 2007). Average commercial landings of red snapper in the Gulf between 2010 and 2012 were worth an estimated $11.7 million, while recreational fishing in 2011 produced an estimated value of $53 million (GMFMC 2013). Red snapper in the Gulf are managed as a single stock that currently is overfished but not undergoing overfishing (SEDAR 2013). The decision to manage the Gulf as a single stock is based, in part, on genetic analyses conducted over the past two decades. These studies (Gold *et al.* 1997, 2001; Pruett *et al.* 2005; Saillant and Gold 2006; Gold and Saillant 2007) were typically based on adult fish with mixed-age classes and generally indicated homogeneity of both nuclear and mitochondrial genetic variation across the northern Gulf.

Alternatively, several lines of evidence suggest that independent stocks may exist in the Gulf of Mexico. Differences in growth rates have been documented between red snapper captured in Louisiana and Alabama as compared to Florida and Texas (Fischer *et al.* 2004; Saari *et al.* 2014) and size and age-at-maturity were found to differ between Louisiana and Alabama (Woods *et al.* 2003). Further, estimates of the effective population size differ significantly between regions (Saillant and Gold 2006; Gold and Saillant 2007) and CPUE trends in the western and eastern Gulf are currently decoupled (Cass-Calay *et* al. 2015). These findings, along with occurrence of differences in habitat type in the east and west (Rezak *et al.* 1985) have led to the proposal of eastern and western sub-stocks occurring on either side of the Mississippi River (SEDAR 2005; Cass-Calay *et al.* 2015), which are not officially used for management but are explicitly modelled for assessment purposes.

Saillant et al. (2010) assessed patterns of spatial and temporal genetic variation among samples of young-of-the year (YOY) red snapper from two cohorts in each of five localities in the northern Gulf, using nuclear-encoded microsatellites. Analysis of molecular variance revealed heterogeneity at small spatial scales, and spatial autocorrelation indicated autocorrelation of genetic variation among fish sampled within 50-100 km. These findings were extended by Puritz *et al.* (2016) in which genomic diversity within and between geographic samples of YOY red snapper were again assessed but this time using 7,382 single nucleotide polymorphisms. Genetic heterogeneity was detected within cohorts but at distances as small as 5 km and no evidence for selection or sweepstake recruitment was found. Taken together these studies suggest that assemblages of YOY red snapper may originate from currently undefined, genetically independent groups of spawners. The idea that marine fishes with long larval periods and large census sizes must have extensive gene flow and completely admixed populations is being challenged by a growing body of research that indicates otherwise (Hauser and Carvallo 2008). The results from Saillant et al. (2010) and Puritz et al. (2016) may reflect a metapopulation-like stock structure in the Gulf of Mexico.

Pruett *et al.* (2005) and Saillant and Gold (2006) previously proposed a metapopulation model following the definition of Kritzer and Sale (2004) where a metapopulation is comprised of a series of partially independent subpopulations that impact one another's demographics

periodically via gene flow and where only a few local or source populations are required to sustain the stock. Consistent with the notion that metapopulation dynamics may be important for red snapper in the Gulf, the geographic distribution of life stages and age classes among red snapper is far from uniform across the northern Gulf (Diamond et al. 2010; Lowerre-Barbieri et al. 2012). Further, significant ontogenetic shifts in habitat usage are well-documented (Szedlmayer and Howe 1997; Szedlmayer and Conti 1999; Gallaway et al. 2009). It may be that distribution patterns of red snapper by life-stage across the Gulf reflect locations of life-stage-specific habitat and in some cases recruitment to these habitats follow a sink-source pattern. For example, replacement of breeding fish along the West Florida Shelf in the absence of small juveniles (Lowerre-Barbieri et al. 2012) might be dependent on west to east movement of post-settlement fish associated with ontogenetic shifts in habitat use, a hypothesis supported by tagging data (Patterson et al. 2001).

The above considerations indicate that rigid, fixed boundaries based on geography may not be a biologically meaningful way to define management units (stocks) of red snapper in the Gulf and that identification of potential 'source' and 'sink' populations should be a priority. Because previous studies utilizing microsatellites and mitochondrial DNA (mtDNA) data were equivocal, we implemented a population genomics approach that allowed us to characterize genetic variation at thousands of polymorphic genetic markers, generated randomly from across the entire genome (Davey and Blaxter 2010), in an attempt to resolve fine-scale patterns of within- and between-group genetic diversity (Allendorf *et al.* 2010). The approach also allows for the identification of genomic regions responding to environmental differences at small spatial scales and underpinning local adaptation (Stapley *et al.* 2010), which may be useful for identifying local management units that differ in aspects of basic biology (Bradbury *et al.* 2010).

To accomplish this, the study was divided into three subprojects. The first subproject involved developing and refining methodologies for using double digest restriction-site associated DNA (ddRAD) sequencing for population genomics of marine fishes. The results of that subproject are detailed in two published manuscripts (Puritz *et al.* 2014; Willis *et al.* 2017) which have been attached in the Supplementary Material. The second subproject, a follow-up to the work of Saillant *et al.* (2010), was focused on assessing patterns of genomic variation among samples of

young-of-the-year red snapper collected throughout the northern Gulf. The results of that subproject are detailed in a published manuscript (Puritz *et al.* 2016) which also is attached in the Supplementary Material. The goal of the third subproject was to assess patterns of genomic variation within and between mixed-age samples of red snapper collected from the U.S. South Atlantic, eastern Gulf, western Gulf and southern Gulf to better understand population structure, connectivity and genetic demography of Gulf red snapper. This third subproject is the focus of the remainder of this report.

III. APPROACH

Sample Collection and Sequencing

Fin clips were obtained from mixed-age samples of red snapper from 11 locations (Figure 1), three in the Atlantic (off North Carolina, NC; off South Carolina, SC; and off Melbourne, Florida, ML) six in the northern Gulf (off the Dry Tortugas, DT; from the Florida Middle Grounds, MG; off Panama City, Florida, PC; from two areas off Louisiana, JC1 and JC2; and from off Texas, TX) and two in the southern Gulf (off Veracruz, VC and from the Campeche Banks, CH). Sampling either involved directed effort, in which case the exact location of catch is known, or port sampling, in which case the exact location of catch is proprietary information of fishers and only the general vicinity is known. All fin clips were preserved in 20% DMSO-0.25M EDTA-saturated NaCl buffer (Seutin *et al.* 1991) or 95% non-denatured ethanol.

DNA was extracted using Mag-Bind Tissue DNA kits (Omega Bio-Tek) and approximately 500 ng of high quality genomic DNA was utilized in a modified version of the ddRAD genomic library preparation method (Peterson *et al.* 2012). In brief, extractions were digested with restriction enzymes *Eco*RI and *Msp*I and a barcoded adapter was ligated to *Eco*RI restriction sites. A common adapter was then ligated to *Msp*I restriction sites, using equimolar quantities of each digested sample. Samples were subsequently pooled into four 'indexed' libraries consisting of ~35 individuals each and size selected using a Pippin Prep DNA size selection system (Sage Science Inc.). Fragments were selected using a mean size of 375 bp, with a selection window of ± 37 bp). Illumina flow-cell adapter sequences and index-specific identifiers were added to each index library, using 12 cycles of PCR. Sequencing was spread out across eight lanes of Illumina HiSeq sequencing with technical replicates (duplicated individuals) sequenced across multiple

lanes.

Bioinformatics and Variant Filtering

Raw Illumina HiSeq reads from each individual were demutliplexed using the "process_radtags" function in the software *Stacks* (Catchen *et al.* 2013). Reads were then trimmed and mapped to a draft assembly of the red snapper genome (154,064 contigs; N50 = 233,156 bp; total length 1.23 Gb) using the *dDocent* v2.18 pipeline (Puritz *et al.* 2014) and contiguous sequence alignments (contigs) approximately 300 base pairs (bp) in length that contained single nucleotide polymorphisms (SNPs) identified. Results were compiled into a variant call file (VCF) file and variants filtered using a combination of VCFTOOLS v0.1.11 (Danacek *et al.* 2011), VCFLIB (E. Garrison, Boston College), and custom BASH and Perl scripts. Contigs, SNPs and individuals were subjected to several iterative filtering steps to improve data quality. First, SNPs that were called in less than 25% of all individuals were removed as well as SNPS with quality scores of less than 10. Then contigs with a minor allele frequency (MAF) of less than 0.5% were removed using a custom script that can handle multialleleic loci. Next individuals with more than 60% missing data were removed from the data set using the script filter_missing_ind.sh (https://github.com/jpuritz/dDocent/blob/master/scripts/filter_missing_ind.sh) and genotypes with less than three reads were changed to missing data. SNPs that were called in less than 75% of individuals were then removed and individuals with more than 27.5% missing data were removed. SNPs were then removed that had more than 20% missing data within any given geographic sampled using the script pop_missing_filter.sh (https://github.com/jpuritz/dDocent/blob/master/scripts/pop_missing_filter.sh) and contigs were then filtered using the script (dDocent_filters; https://github.com/jpuritz/dDocent/blob/master/scripts/dDocent_filters) as described in Puritz *et al.* (2016) with a maximum mean depth cutoff of 145. Technical replicates were then screened for non-matching genotypes at each SNP, and these SNPs were removed from the data set. Variant calls were then broken into SNP and insertion/deletion mutations (INDELs), using vcfallelicprimatives from the package vcflib (https://github.com/vcflib/vcflib); INDELs were removed using VCFtools to produce a VCF file of contigs containing only SNPs. SNP calls were then filtered to only include alleles with a MAF > 5%. Finally, problematic and biased contigs, identified as described in Puritz *et al.* (2016), as well as by multiple rounds of outlier detection

and visual alignment inspection, were removed to produce a final vetted data set of SNPs. The data set was then phased into constituent haplotypes using a custom Perl script following Willis *et al.* (2017), creating a final data set of multi-allelic SNP-containing loci (hereafter loci).

Genomic diversity and outlier detection

Rarefied allelic richness (A_R, El Mousadik and Petit 1996) and gene diversity (H_E, Nei 1973) at each locality were estimated using HEIRFSTAT V 0.0.48 (Goudet 2005) and ARLEQUIN V 3.5.1.2 (Excoffier and Lischer 2010), respectively. The data set was screened for outlier loci using the Bayesian modeling approach implemented in BAYESCAN v2.1 (Foll and Gaggiotti, 2008). Analysis involved 30 pilot runs of 5000 steps, a burn-in of 50,000 steps, and 500,000 MCMC samples with a thinning interval of 100; significance was determined using a q-value = 0.05. Outlier loci are potentially under positive or diversifying selection and can provide evidence of localized selective pressure and adaptations. However, they also may provide misleading signal with regards to genetic demography which is better addressed using only loci presumed to be selectively neutral (Funk *et al.* 2012). Therefore, identified outlier loci were separated from putatively neutral loci to create a neutral and non-neutral (outlier) data set and downstream data analyses were conducted separately for both data sets, unless otherwise noted.

Population genetic analyses

Exact test of genic and genotypic differentiation were run in GENEPOP V.4.7.0 (Raymond and Rousset 1995; Rousset, 2008). Significance was assessed, with 10,000 dememorizations, 100 batches and 5,000 iterations per batch. A single-level, locus-by-locus AMOVA was performed using ARLEQUIN, with global F_{ST} calculated as a weighted mean of locus-specific F_{ST} values to account for uneven levels of missing data across loci (Weir and Cockerham 1984). Significance was assessed by permuting individuals among samples 10,000 times. Pairwise F_{ST} also was calculated using a locus-by-locus framework in ARLEQUIN. Significance was again assessed by permuting individuals between samples 10,000 times with significance corrected for multiple comparisons using the Benjamini and Hochberg (1995) false discovery rate procedure.

Spatial analysis of principle components (sPCA) was run using the R-package (R Core Team 2017) adegenet 2.0.1 (Jombart 2008; Jombart and Ahmend 2011). This approach uses multi-

locus, individual-based genotypic data to form groups where within-group differences are minimized and between-group differences are maximized, allowing for an assessment of optimal genetic groupings without an *a priori* hypothesis. Further, the approach does not require that populations are in equilibrium, an assumption that is likely violated in many marine species (Portnoy and Gold 2013). Finally, sPCA takes spatial autocorrelation into account and may therefore outperform analyses like PCA in detected cryptic patterns of population structure (Jombart 2008). Analyses were run twice, once using a nearest neighbor network (Figure 2a) and once using a Delaunay triangulation (Figure 2b), to assess the effect that network choice had on clustering. A DAPC was then run with individuals placed into groups (regions) defined by sPCA results and pairwise F_{ST} between regions estimated in a locus by locus framework.

The contemporary effective size (N_E) was estimated for each region defined by sPCA, using only the neutral data set and the linkage disequilibrium approach of LDNE (Waples 2006; Waples and Do 2008), as implemented in NEESTIMATOR 2 (Do *et al.* 2014). Minor alleles at a frequency of 0.02 or less were excluded, following Portnoy *et al.* (2009), and confidence intervals were obtained parametrically. A global estimate of contemporary effective size (N_E) was also obtained by pooling all samples.

Connectivity

Patterns of contemporary connectivity were assessed in two ways. First, to assess whether an individual was a first generation migrant, likelihood ratios were calculated as the likelihood of an individual's multilocus genotype originating in the region from which the individual was sampled over the highest likelihood of that genotype originating across all regions ($L' = L_{home}/L_{max}$) using a Bayesian approach (Rannala and Mountain 1997) as implemented in GeneClass 2.0.h (Piry *et al.* 2004). Significance was determined by simulating 10,000 individuals (Paetkau *et al.* 2004) and generating a null distribution of L' values, observed L' values were evaluated at $\alpha \leq 0.01$. Second, contemporary migration rates (m) between regions were estimated using the Bayesian approach implemented in BiMR v1.0 (Faubet and Gaggiotti 2008). Analysis involved 40 pilot runs of 10,000 steps, followed by 10 full runs each with a burn-in of 500,0000 steps, and 1,000,000 MCMC samples with a thinning interval of 100. Final estimates of m were made from the run with the lowest Bayesian deviance for assignment

(Faubet *et al.* 2007) and 95% highest posterior density intervals (HPDI) were evaluated to look for statistically significant differences in *m* between pairs of samples (migration asymmetry).

IV. FINDINGS

Genomic diversity and outlier detection

A total of 2,138,152 SNP variants were mapped to the red snapper draft genome and, after filtering the final data set contained 5,591 SNPs in 1,822 loci genotyped for 318 individuals. There were between 1 and 10 SNPs and 2 and 32 haplotypes per locus. Estimates of average A_R across loci range from 2.495 at VC (±SE; 0.024) to 2.515 at DT (± 0.024) and estimate of average H_E across loci ranged from 0.391 at DT (± 0.005) to 0.385 at LA1 (± 0.005). Estimates of within sample diversity are shown in Table 1. Eighteen significant outlier loci were detected at $q \leq 0.05$.

Population genetic analyses

Exact test of global genic and genotypic differentiation were significant for both the neutral ($P < 0.0007$) and outlier data set ($P < 0.0001$). Results of the single-level, locus-by-locus AMOVA indicated a significant component of variance was attributable to differences among geographic samples for both the neutral ($P = 0.016$) and outlier ($P < 0.0001$) data sets (Table 2). For the neutral dataset ten estimates of pairwise F_{ST}, out of 55 total estimates, were significant before correction for multiple comparisons, none were significant after correction (Supplemental Table 1). For the outlier dataset 46 estimates of pairwise F_{ST}, out of 55 total estimates, were significant before correction for multiple comparisons, 44 were significant after correction (Supplemental Table 2).

Spatial analysis of principle components using the neutral data set identified different clustering of samples depending on which connection network was used to define the spatial relationships between the samples. When using a nearest neighbor network the samples formed four clusters (Figure 3a); one contained the Mexican samples (hereafter SGulf), one contained the samples from Louisiana and Texas (hereafter WGulf), one contained all samples from off the coast of Florida (hereafter FL), and one containing the two samples from the Carolinas (hereafter Car). When Delaunay triangulation was used to define the connection network the samples formed two

clusters (Figure 3b), one cluster contained the Florida and Carolina samples (hereafter FL-Car) and the other contained the Mexican samples, TX and LA (hereafter SWGulf); there was overlap with SC, DT, ML and CH along the first two principal components. For the outlier data (Figure 4) similar results were obtained regardless of network, with all samples clustered and TX and LA slightly separated in one direction along the first axis and ML slightly separated in the other direction. For the remainder of analyses results from the neutral data were considered as potential models of population structure and are called the four-region model (SGulf, WGulf, FL and Car) and two-region model (SWGulf and FL-Car).

When DAPC was run using neutral data and the four-region model, the regions showed separation with FL intermediate to the other three regions (Figure 5a). By contrast, when DAPC was run using outlier data and the four-region model, the regions showed greater separation but Car was intermediate to the other three regions (Figure 5b). Estimates of pairwise F_{ST} between regions were not significant before or after correction for the neutral data set but were significant for four out of six comparisons after correction for the outlier data set (Table 3). DAPC was not run using the two-region model but pairwise F_{ST} between regions was significant but again only for the outlier data set (Table 3).

Point estimates of N_E for the regions defined by the four-region model were infinite, except for FL which had a point estimate of 64,934, and lower bounds were always greater than 19,500 (Table 4). By contrast, point estimates of N_E for both regions defined by the two-region model were ~ 40,000 and had confidence intervals with finite upper bounds. The global estimate of N_E was 65,733 and associated confidence intervals had finite upper bound.

Connectivity

Detection of contemporary migration requires that populations are genetically distinguishable (Paetkau *et al.* 2004; Faubet *et al.* 2007), therefore, only the outlier data set was used for migrant detection and estimation of m. Four individuals were flagged as potential migrants using the four-region model ($p < 0.01$, Table 5) and nine were individuals were flagged as potential migrants using the two-region model ($p < 0.01$, Table 5). Only two individuals were flagged in common across the four-region and two-region model; both were caught in Florida and their

most likely population of origin was SGulf (SWGulf). Point estimates of *m* for the four-region model were between 5-7% for migration into Florida, while migration into all other regions was negligible (Table 6). Similarly, for the two-region model migration going into FL-CAR was much higher than in the other direction (Table 6).

V. EVALUATION

The results of this subproject indicate that red snapper are genetically heterogenous across the current spatial sampling. While estimates of pairwise F_{ST} for all samples (using neutral and outlier loci) were difficult to interpret, analyses (sPCA) that accounted for spatial autocorrelation of genetic variation and do not depend on equilibrium assumptions were suggestive of either two groups, eastern Gulf-Atlantic and a western Gulf-southern Gulf, or four groups, Carolinas, Florida, western Gulf and southern Gulf. Estimates of contemporary effective size for these regions, when finite, were similar in magnitude and large (>40,000) as was the global estimate (65,733) and estimates of contemporary migration were generally low between the putative regions, with higher estimated migration rates into Florida. The data provided here add to a growing body of research that supports the idea that red snapper in U.S. waters are not one well-mixed stock but instead multiple independent reproductive units that may interact as a metapopulation due to varying levels of connectivity (Pruett *et al.* 2005).

Genetic analysis of stock structure in red snapper has been an ongoing field of investigation for 20 years and the current view of stock structure results from a synthesis of these studies. While, initial analyses, using mtDNA restriction length fragment polymorphisms and *F*-statistics, found no strong evidence of structure among localities sampled in the northern Gulf of Mexico; differences in within sample diversity were consistent with the possibility of independent populations with recent co-ancestry (Gold *et al.* 1997). A follow up study, utilizing 20 nuclear-encoded microsatellite loci, again found little evidence of population structure from *F*-statistics, though one locus indicated spatial heterogeneity (Gold *et al.* 2001). Pruett *et al.* (2005), using mtDNA sequence data, detected patterns of genetic variation consistent with isolation by distance and population expansion, but again failed to detect significant spatial heterogeneity using traditional *F*-statistics. The authors suggested that the result might reflect metapopulation

structure, where independent demes were connected by temporally heterogenous levels of gene flow. Saillant and Gold (2006) and Gold and Saillant (2007) provided further evidence for metapopulation structure when they showed that significant differences in microsatellite allele frequencies existed across spatially discrete samples of adults within cohorts, but that these differences were not consistent across cohorts. Similarly, Saillant *et al.* (2010) found genetically divergent groups of YOY across the northern Gulf of Mexico, but was unable to discern whether the pattern was attributable to independent groups of spawning adults or sweepstakes recruitment. By using a genomics approach Puritz *et al.* (2016) demonstrated that the genetic differences among spatially discrete groups of YOY was unlikely to be the result of either sweepstakes recruitment or localized selection and instead likely reflected metapopulation dynamics originally described by Pruett *et al.* (2005).

The results of subproject three are similar in that differences between regions were detected by methodologies that do not rely on equilibrium assumptions (Pruett *et al.* 2005; Hollenbeck *et al.* 2015), but not by traditional *F*-statistics. *F*-statistics rely on assumptions of equilibrium between microevolutionary processes (i.e. drift, gene flow, mutation, selection) and those assumptions are likely to be violated if populations have experienced recent divergence/expansion and/or fluctuating level of connectivity (Portnoy and Gold 2013) as has been suggested for Gulf red snapper (Pruett *et al.* 2005). Unlike previous studies, the results here suggest that there may be western and eastern Gulf groups, similar to the management subunits defined by SEDAR (2009), but with the eastern Gulf group inclusive of red snapper off the Atlantic coast of Florida. The east west groupings also are consistent with life history data that suggests regional differences in size-at-age and growth in the Gulf of Mexico (Fischer *et al.* 2004; Saari *et al.* 2014) and a recent observed decoupling of CPUE data between the eastern and western Gulf (Cass-Calay *et al.* 2015).

The genomics approach used in this project also allowed for the detection of F_{ST} outlier loci, which may represent regions of the genome under directional selection (Nielson *et al.* 2009; Allendorf *et al.* 2010) or regions that have diverged more than expected over time through non-adaptive processes such as genetic drift and isolation (Hedrick 2011). In the context of this study, however, the outlier loci were important because they provided enough discriminatory power

between regions to allow for an assessment of relative magnitude of contemporary migration. Though estimated migration rates differed between methodological approaches and the two or four-region model, there were some important consistencies among analyses. First, the estimated migration rates between regions was generally small, less than 10%, meaning that regions are likely to respond independently to perturbations (Hastings 1993; Hauser and Carvalho 2008; Waples 2010). Second, migration asymmetry generally favored movement from the west and south toward the east and north, and more specifically in the four-region model, into Florida. This result is highly congruent with tagging data that have demonstrated eastward movement of red snapper in the northern Gulf (Patterson *et al.* 2001) and observed differences in distribution of age classes and life stages across the Gulf (Lowerre-Barbierri *et al.* 2012; Karnauskas *et al.* 2017). This supports the notion that red snapper biomass in the eastern Gulf (Florida) may be supplemented through migration from the west.

While this subproject provides support for regional structuring of red snapper in U.S. waters, it is important to note that the results are suggestive rather than conclusive. Sampling was largely opportunistic and thus there are differences in the age-composition of samples. Further, when port sampling was used, information about specific location of capture and sometimes size data were not available. This means that failure to find consistent results across analyses may in part be due to violations of assumptions implicit in certain analysis (i.e. equilibrium) but also may be in part due to sampling itself. While there is no question that red snapper do not form a single homogenous gene pool across the Gulf of Mexico, more work will be required to accurately define the spatial extent/location of independent groups of spawners. Ultimately, a sampling strategy that is representative of age structure and differential habitat use may be required and access to samples from actively spawning adults across multiple years would likely be needed as well to accurately describe red snapper population structuring.

VI. Tables

Table 1: Number of individuals genotyped (N) and estimates of rarified allelic richness (A_R) and gene diversity (H_E) averaged across loci, with standard error (+/-), for each geographic samples of red snapper**.**

	N	A_R	+/-	H_E	+/-
NC	29	2.499	0.024	0.386	0.005
SC	31	2.499	0.024	0.386	0.005
ML	21	2.492	0.024	0.386	0.005
DT	29	2.515	0.024	0.391	0.005
MG	20	2.494	0.025	0.386	0.005
PC	30	2.512	0.024	0.390	0.005
JC1	38	2.497	0.024	0.385	0.005
JC2	35	2.505	0.024	0.386	0.005
TX	23	2.500	0.024	0.387	0.005
VC	30	2.495	0.024	0.387	0.005
CH	32	2.501	0.024	0.387	0.005

Table 2: Results of single level AMOVA for both the neutral and outlier data sets; SS, sum of squares; VC, variance component; and %V, percentage of variance

	SS	VC	% Variance	F_{ST}	P-value
Neutral loci					
Among populations	3866.78	0.69	0.20	0.002	0.016
Within populations	210971.41	348.62	99.80		
Total	214838.19	349.31			
Outlier loci					
Among populations	109.25	0.13	3.24	0.032	<0.0001
Within populations	2265.59	3.91	96.76		
Total	2374.84	4.04			

Table 3: Estimates of pairwise F_{ST} (above the diagonal) for regions defined by the four-region and two-region models, based on the neutral and outlier data sets. *p*-values are displayed below the diagonal and bolded values are significant after correction.

	Outliers				**Neutral**			
	SWGulf		FL_CAR		SWGulf		FL_Car	
SWGulf	*		**0.013**		*		0.00035	
FL_Car	**> 0.001**		*		0.258		*	
	SGulf	WGulf	FL	Car	SGulf	WGulf	FL	Car
SGulf	*	**0.027**	0.012	0.004	*	0.00083	0.00065	0.00064
WGulf	**> 0.001**	*	**0.030**	**0.022**	0.169	*	0.00064	0.00092
FL	**> 0.001**	**> 0.001**	*	0.005	0.352	0.184	*	0.00064
Car	0.191	**> 0.001**	0.083	*	0.607	0.115	0.431	*

Table 4: Estimates of contemporary effective size (N_E) for regions defined by the four-region and two-region models, made using the linkage disequilibrium approach; confidence intervals (CI) calculate parametrically.

	N_E	95% CI	
Regions	**Point**	**Lower**	**Upper**
Car	Infinite	19,687	Infinite
FL	64,934	20,291	Infinite
WGulf	Infinite	Infinite	Infinite
SGulf	Infinite	29,485	Infinite
FL-Car	40,841	22,292	241,424
SWGulf	45,396	23,656	551,059
All	65,733	39,796	188,200

Table 5: Numbers of real-time potential migrants between regions estimated using outlier data, for the two-region and four-region models, identified using likelihood ratio tests (p <0.01). Region in which individual was captured along the top, region with highest likelihood of origin along the side.

	SWGulf	FL_Car
SWGulf	-	**5**
FL_Car	**4**	-

	S Gulf	W Gulf	FL	CAR
S Gulf	-	0	**2**	0
W Gulf	0	-	0	0
FL	**1**	0	-	**1**
Car	0	0	0	-

Table 6: Contemporary migration rate (*m*) between regions, for the four-region and two-region models, estimated in a Bayesian framework using outlier data, with 95% highest posterior density intervals (HPDI).

From	Into	Mode	Low	High
WGulf	SGulf	3.24E-19	3.82E-06	0.000107
FL	SGulf	1.64E-20	2.81E-07	1.39E-05
Car	SGulf	4.09E-20	3.05E-06	6.48E-05
SGulf	WGulf	8.71E-19	8.71E-19	6.54E-05
FL	WGulf	1.38E-19	5.33E-06	0.000221
Car	WGulf	5.33E-19	4.88E-07	4.80E-05
SGulf	FL	0.0655	0.0113	0.301
WGulf	FL	0.0517	0.0115	0.291
Car	FL	0.0697	0.0153	0.312
SGulf	Car	1.64E-18	7.07E-07	3.89E-05
WGulf	Car	8.28E-18	3.41E-07	2.11E-05
FL	Car	6.71E-19	5.14E-07	5.05E-05
SWGulf	FL_Car	0.412	0.112	0.834
FL_Car	SWGulf	0.0872	0.00723	0.334

VII. Figures

Figure 1: Approximate sampling locations associated with 11 geographic samples included in the current study; WL (North Carolina; N= 29), SC (South Carolina, N= 31), ML (Melbourne FL, N=21), DT (Dry Tortugas, N=29), MG (Middle Grounds off Florida, N=20), PC (Panama City and vicinity, N=30), JC2 (off Louisiana, N= 35), JC1 (off Louisiana, N=38), TX (off Texas, N=23), VC (off Vera Cruz, Mexico, N=30) and CH (Campeche Banks, Mexico, N=32).

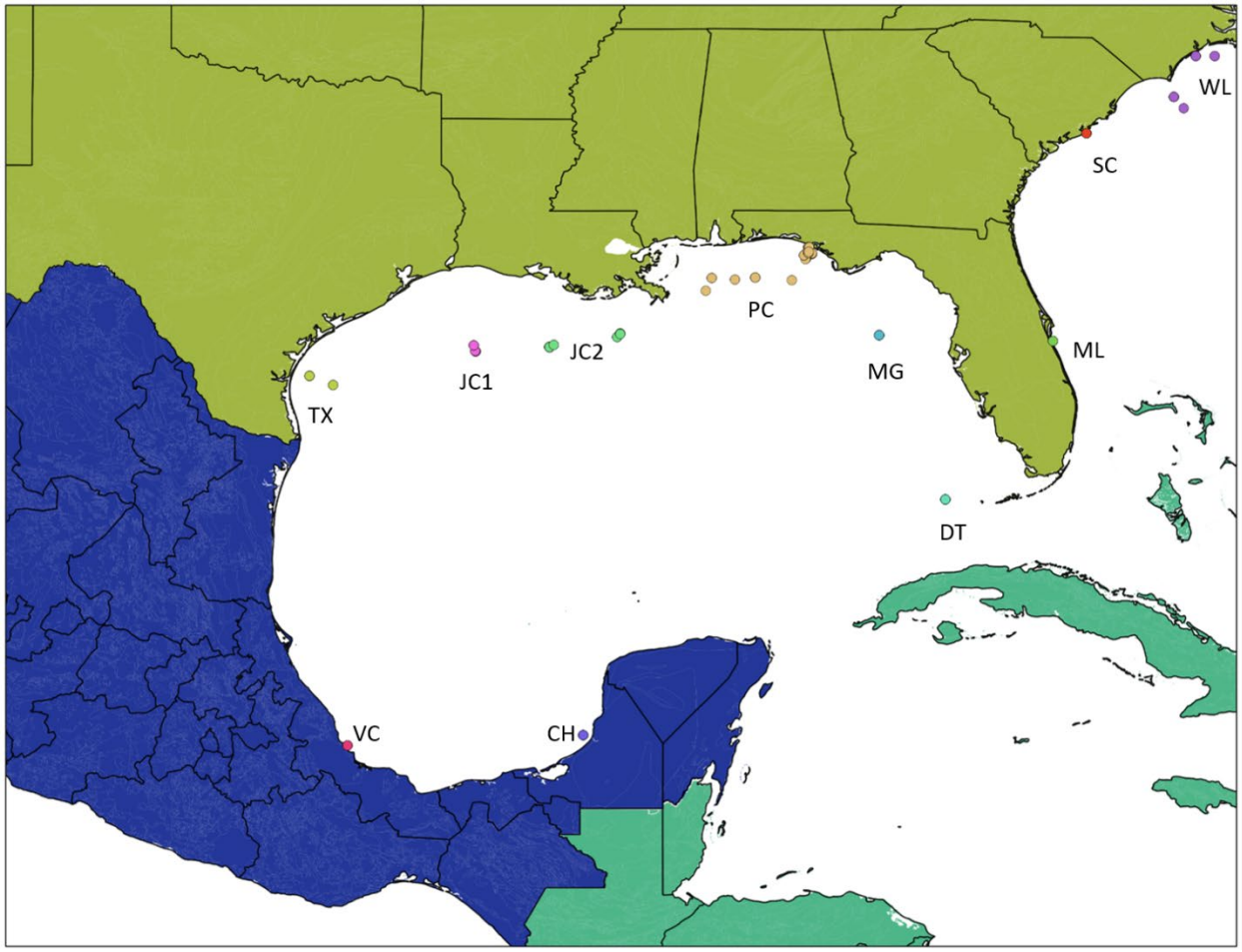


Figure 2: Visual representation of connection networks used for spatial analysis of principal components (sPCA); nearest neighbor network (a) and Delaunay triangulation (b).

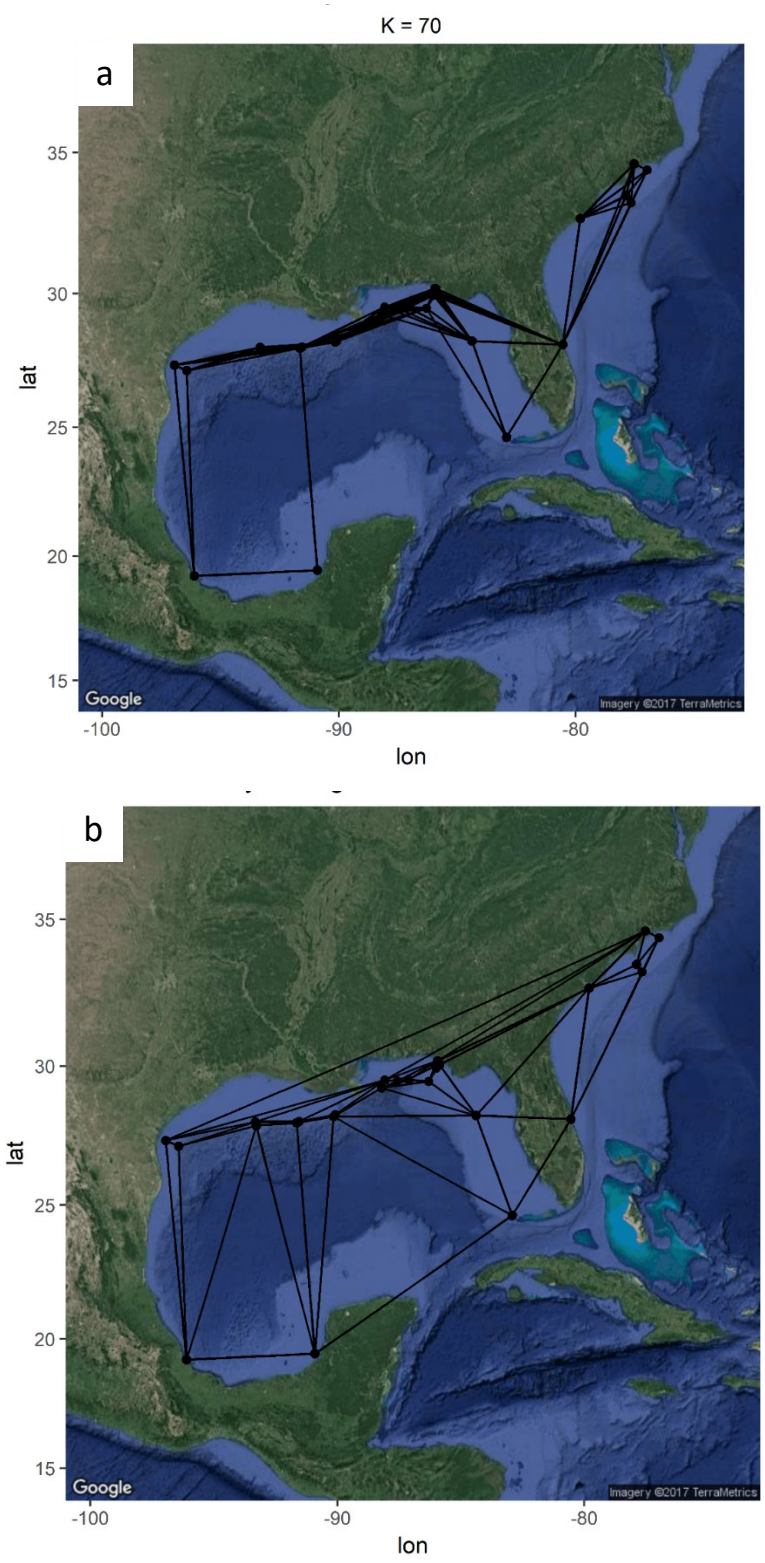


Figure 3: Visualization of spatial analysis of principal components (sPCA) as a contour plot and standard PCA plot using the neutral data set; nearest neighbor network (a) and Delaunay triangulation (b).

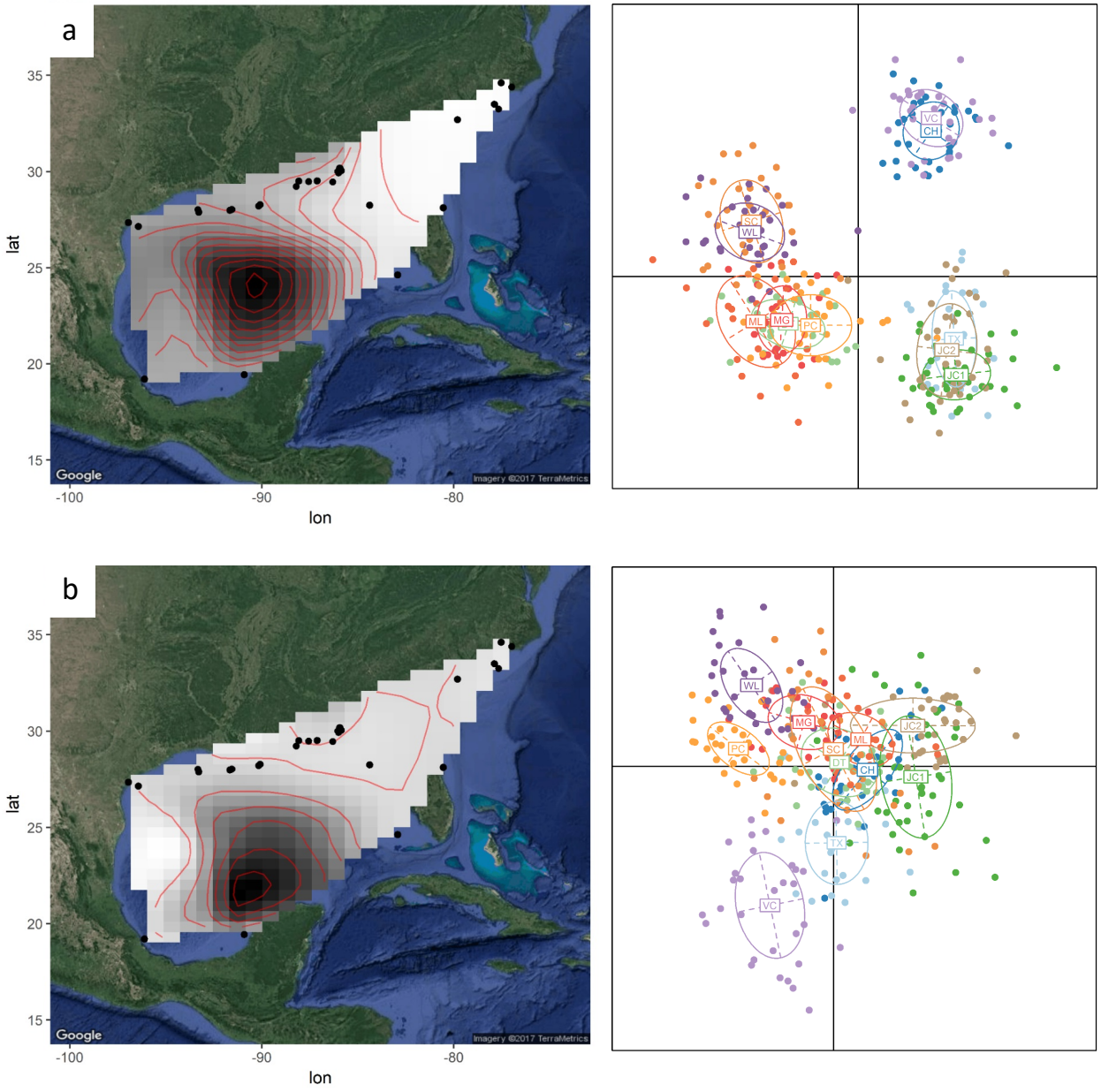


Figure 4: Visualization of spatial analysis of principal components (sPCA) as a contour plot and standard PCA plot using the outlier data set; nearest neighbor network (a) and Delaunay triangulation (b).

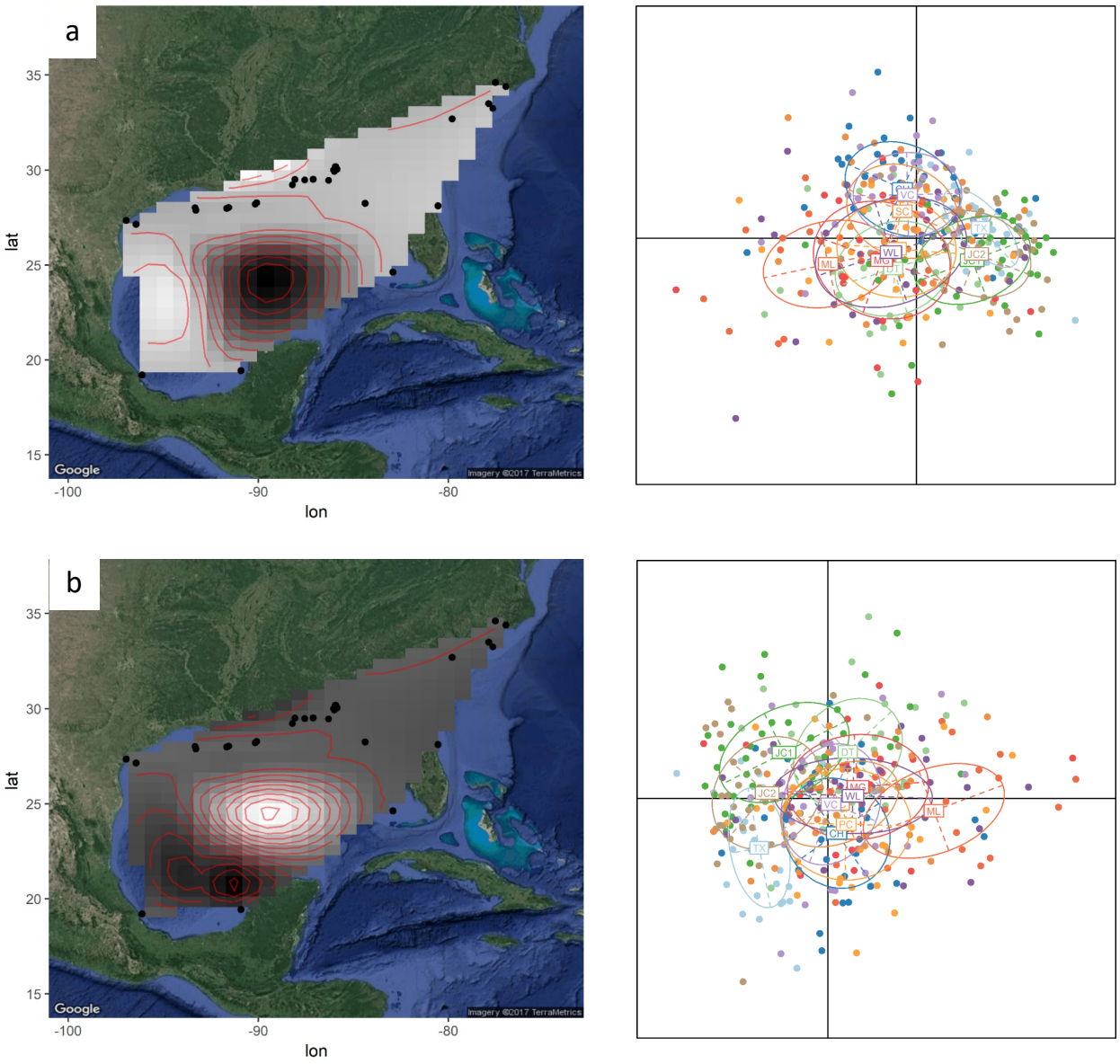
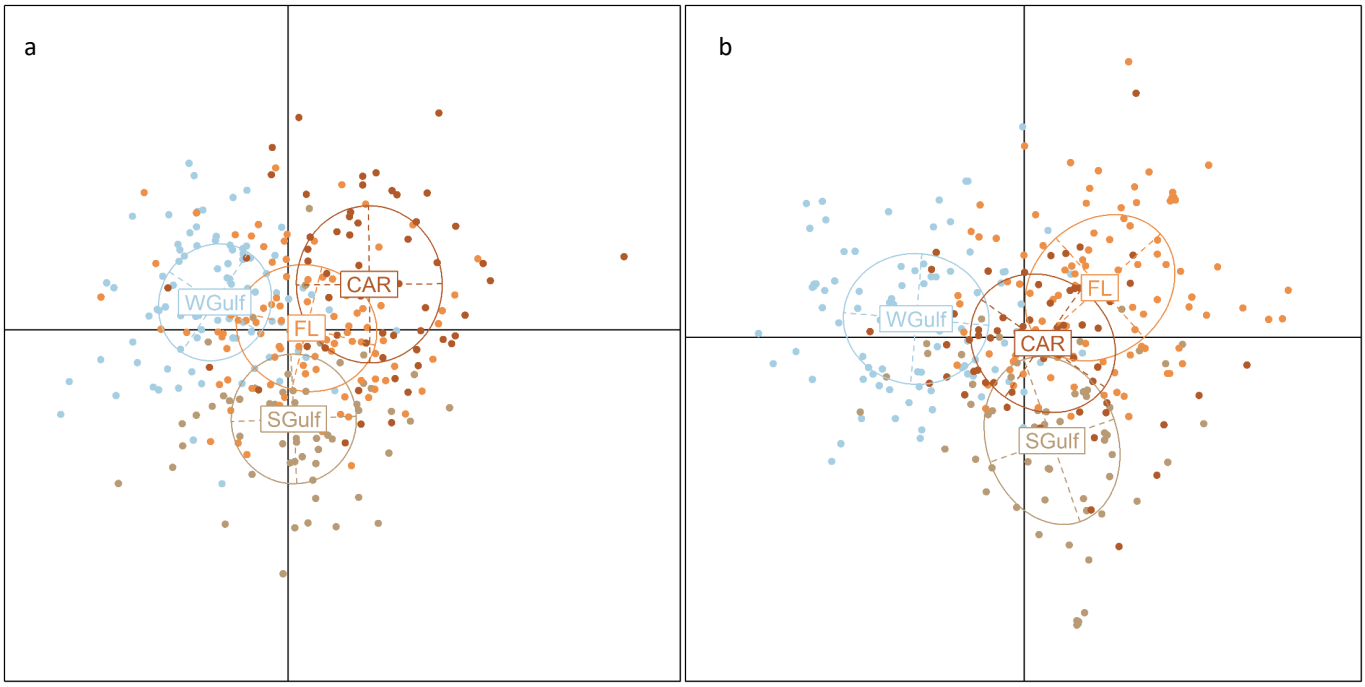


Figure 5: Discriminant analysis of principle components (DAPC) using predefined membership based on the four-region model for the neutral data set (a) and the outlier data set (b).



IX. References Cited

Allendorf, F. W., Hohenlohe, P. A., Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews: Genetics*, 11: 697–709.

Benjamini, Y., Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B (Statistical Methodology)*, 57: 289–300.

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., Bentzen, P. (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society of London B: Biological Sciences B*, 277: 3725-3737.

Cass-Calay, S. L., Porch, C. E., Goethel, D. R., Smith, M. W., Matter, V., McCarthy, K. J. (2015) Stock assessment of red snapper in the Gulf of Mexico 1872-2013 with provisional 2014 landings. SEDAR 31 Gulf of Red Snapper Update Report. North Charleston, SC.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., Cresko, W. A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22: 3124-3140.

Danecek, P., Auton, A., Abecasis, G., *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, 27: 2156-2158.

Davey, J. L., Blaxter, M. W. (2010) RADseq: Next-generation population genetics. *Briefings in Functional Genomics*, 9: 416–423.

Diamond, S. L., Kleisner, K. M., Duursmae, D. E., Wang, Y. (2010) Designing marine reserves to reduce bycatch of mobile species: A case study using juvenile red snapper (*Lutjanus campechanus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 67: 1335-1349.

Do C., Waples R. S., Peel D., Macbeth G. M., Tillett B. J., Ovenden J. R. (2014) NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14: 209-214.

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

Excoffier L., Lischer H. E. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10: 564-567.

Faubet, P., Waples, R. S., Gaggiotti, O. E. (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, 16: 1149 1166.

Faubet P. and O.E. Gaggiotti, 2008. A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics*, 178: 1491–1504.

Fischer, A. J., Baker Jr., M. S., Wilson, C. A. (2004) Red snapper *(Lutjanus campechanus*) demographic structure in the northern Gulf of Mexico based on spatial patterns in growth rates and morphometrics. *Fisheries Bulletin*, 102: 593–603.

Foll, M., Gaggiotti, O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180: 977-993.

Funk, W. C., McKay, J. K., Hohenlohe P. A., Allendorf F. W. (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27: 489-496.

Gallaway, B. J., Szedlmayer, S. T., Gazey, W. J. (2009) A life history review for red snapper in the Gulf of Mexico with an evaluation of the importance of offshore petroleum platforms and other artificial reefs. *Reviews in Fisheries Science*, 17: 48-67.

GMFMC (2013) Red Snapper 2013 Quota Increase and Supplemental Recreational Season Including Environmental Assessment, Regulatory Impact Review, and Regulatory Flexibility Act Analysis. Gulf of Mexico Fishery Management Council. Tampa, FL.

Gold, J. R., Sun, F., Richardson, L. R. (1997) Population structure of red snapper from the Gulf of Mexico as inferred from analysis of mitochondrial DNA. *Transactions of the American Fisheries Society*, 126: 386–396.

Gold, J. R., Pak, E., Richardson, L. R. (2001) Microsatellite variation among red snapper (*Lutjanus campechanus*) from the Gulf of Mexico. *Marine Biotechnology*, 3: 293–304.

Gold, J. R., Saillant, E. (2007) Population structure of red snapper (*Lutjanus campechanus*) from the Northern Gulf of Mexico. In: Patterson III, W. F., Cowan Jr., J. H., Fitzhugh, G. R., Nieland, D. L. (Eds.), *Red Snapper Ecology and Fisheries in the U.S. Gulf of Mexico*. American Fisheries Society, Bethesda, MD, pp 201-216.

Goudet, J. (2005) Hierfstat, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes,* 5: 184-186.

Hastings, A. (1993) Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. *Ecology*, 74: 1362–1372.

Hauser, L., Carvalho, G. C. (2008) Paradigm shifts in marine fisheries genetics ugly hypothesis slain by beautiful facts. *Fish and Fisheries*, 9: 333-362.

Hedrick, P. W. (2011) *Genetics of Populations, 4th Edition*. Jones and Bartlett.

Hollenbeck, C. M., Portnoy, D. S., Saillant, E., Gold, J. R. (2015) Population structure of red snapper (*Lutjanus campechanus*) in US waters of the western Atlantic Ocean and the northeastern Gulf of Mexico. *Fisheries Research*. 172: 17-25.

Hood, P., Strelcheck, A. (2007) A history of red snapper management in the Gulf of Mexico. In: Patterson III, W.F., Cowan Jr., J.H., Fitzhugh, G.R., Nieland, D.L.(Eds.), *Red Snapper Ecology and Fisheries in the U.S. Gulf of Mexico*. American Fisheries Society, Bethesda, MD, pp. 241–256.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24: 1403-1405.

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

Karnauskas, M., Walter III, J. F., Campbell, M. D., Pollack, A. G., Drymon, J. M., Powers, S. (2017) Red Snapper Distribution on Natural Habitats and Artificial Structures in the Northern Gulf of Mexico. *Marine and Coastal Fisheries*, 9: 50-67.

Kritzer, J. P., Sale, P. F. (2004) Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. *Fish and Fisheries*, 5: 131-140.

Lowerre-Barbieri, S. L., Crabtree, T. S., Switzer, T. S., McMichael, R. H. (2012) Spatio-temporal dynamics in red snapper reproduction on the West Florida Shelf, 2008-2011. SEDAR 31-DW15. North Charleston, SC.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Science USA*, 70: 3321-3323.

Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F. & Bekkevold, D. (2009) Population genomics of marine fishes: Identifying adaptive variation in space and time. *Molecular Ecology*, 18: 3128–3150.

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.

Patterson III, W. F., Watterson, J. C., Shipp, R. L., Cowan Jr., J. H. (2001) Movement of Tagged Red Snapper in the Northern Gulf of Mexico. *Transactions of the American Fisheries Society*, 130: 533-545.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., Hoekstra, H. E. (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**: e37135.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., Estoup, A. (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity,* **95**: 536–539.

Portnoy, D. S., McDowell, J. R., McCandless, C. T., Musick, J. A., Graves, J. E. (2009) Effective size closely approximates the census size in the heavily exploited western Atlantic population of sandbar sharks, *Carcharhinus plumbeus. Conservation Genetics*, 10: 1697-1705.

Portnoy, D. S., Gold, J. R. (2013) Finding geographic population structure in marine fish species with high gene flow. *Gulf and Caribbean Fisheries Institute* 65, 384-389.

Pruett, C.L., Saillant, E., Gold, J.R. (2005) Historical population demography of red snapper (*Lutjanus campechanus*) from the northern Gulf of Mexico based on analysis of sequences of mitochondrial DNA. *Marine Biology*, 147: 593-602.

Puritz, J. B., Hollenbeck, C. M., Gold, J. R. (2014) *dDocent*: a RADseq, varian-calling pipeline designed for population genomics of non-model organisms. *PeerJ,* 10: e431

Puritz, J. B., Gold, J. R., Portnoy, D. S. (2016) Fine-scale partitioning of genomic variation among recruits in an exploited fishery: causes and consequences. *Nature: Scientific Reports*, 6: 36095.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rannala, B., and Mountain, J. L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Science, USA*, 94: 9197–9201.

Raymond, M., Rousset, F. (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86: 248-249.

Rezak, R., Bright, T. J., McGrail, D. W. (1985). *Reefs and banks of the northwestern Gulf of Mexico: their geological, biological and physical dynamics*. Wiley, New York.

Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8: 103-106.

Saari, C. R., Cowan Jr., J. H., Boswell, K.M. (2014) Regional differences in the age and growth of red snapper (*Lutjanus campechanus*) in the U.S. Gulf of Mexico. *Fisheries Bulletin*, 112: 261-273.

Saillant, E., Gold, J. R. (2006) Population structure and variance effective size of red snapper (*Lutjanus campechanus*) in the northern Gulf of Mexico. *Fisheires Bulletin*, 104: 136–148.

Saillant, E., Bradfield, S.C., Gold, J.R. (2010). Genetic variation and spatial autocorrelaation among youbg-of-the-yrar red snapper (*Lutjanus camepchanus*) in the northern Gulf of Mexico. *ICES Journal of Marine Science*, 67, 1240-1250.

SEDAR (2005) Southeast Data, Assessment, and Review: SEDAR07 Gulf of Mexico Red Snapper Stock Assessment Report. Charleston, SC.

SEDAR (2009) Southeast Data, Assessment, and Review: 2009 Update, SEDAR07 Gulf of Mexico Red Snapper Stock Assessment Report. Charleston, SC.

SEDAR (2013) Southeast Data, Assessment, and Review: SEDAR 31 Gulf of Mexico Red Snapper Stock Assessment Report. Charleston, SC.

Seutin, G., White, B. N., Boag, P. T. (1991) Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, 69: 82–90.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball A. D., Beckerman, A. P., Slate J. (2010) Adaptation genomics: The next generation. *Trends in Ecology and Evolution,* 25: 705–712.

Szedlmayer, S. T., Howe, J. C. (1997) Substrate preference studies in age-0 red snapper, *Lutjanus campechanus*. *Environmental Biology of Fishes*, 50: 203-207.

Szedlmayer, S. T., Conti, J. (1999) Nursery habitats, growth rates, and seasonality of age 0 red snapper, *Lutjanus campechanus*, in the northeast Gulf of Mexico. *Fisheries Bulletin*, 97: 626–635.

Waples, R. S. (2006) A bias correction for estimates of effective population size based on

linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, 7: 167-184.

Waples, R. S. (2010) Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. Molecular Ecology Resources, 10: 785–796.

Waples, R. S., Do, C. (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, 8: 753-756.

Weir, B. S., Cockerham, C. C. (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38: 1358-1370.

Willis, S. C., Hollenbeck C. M., Puritz, J. B., Gold, J. R., Portnoy D. S. (2017) Haplotyping RAD loci: an efficient method to filter multi-copy loci and account for linkage disequilibrium. *Molecular Ecology Resources*, 17: 955-965.

Woods, M. K., Fischer, A. J., Cowan Jr, J. H., Nieland, D. L. (2003) Size and age at maturity of female red snapper *Lutjanus campechanus* in the northern Gulf of Mexico. *Proceeding of the Gulf and Caribbean Fisheries Institute*, 54: 526-537.

X. Supplemental Materials

Table S1: Estimates of pairwise F_{ST} (above the diagonal) using the neutral data set with associated *p*-values (below the diagonal); bolded estimates are significant before correction.

	WL	SC	ML	DT	MG	PC	JC2	JC1	TX	VC	CH
WL	*	**0.00249**	0.00256	0.00234	0.00057	0.00114	0.00198	0.00230	0.00296	0.00182	0.00191
SC	**0.03960**	*	**0.00311**	0.00209	0.00143	0.00230	0.00226	**0.00216**	**0.00285**	0.00204	0.00111
ML	0.16099	**0.01149**	*	0.00271	0.00289	0.00268	**0.00309**	0.00168	0.00285	**0.00368**	**0.00295**
DT	0.09535	0.08624	0.06000	*	0.00115	0.00144	0.00196	0.00185	0.00223	0.00166	0.00120
MG	0.99168	0.73446	0.20188	0.86762	*	0.00019	0.00130	0.00149	0.00273	0.00139	0.00036
PC	0.86436	0.06505	0.11574	0.57238	0.99733	*	0.00244	0.00221	0.00186	0.00143	0.00177
JC2	0.30871	0.06168	**0.03762**	0.21653	0.91168	0.07218	*	0.00191	0.00258	**0.00260**	0.00139
JC1	0.06426	**0.04327**	0.51396	0.17158	0.75931	0.07723	0.17673	*	**0.00297**	**0.00233**	0.00143
TX	0.05772	**0.02644**	0.11970	0.19901	0.31059	0.51990	0.12812	**0.01337**	*	0.00226	0.00258
VC	0.38168	0.12228	**0.00267**	0.36297	0.81851	0.62188	**0.02723**	**0.03455**	0.23069	*	0.00170
CH	0.20446	0.61723	**0.01495**	0.57832	0.98584	0.26653	0.53762	0.37317	0.05158	0.25535	*

Table S2: Estimates of pairwise F_{ST} (above the diagonal) using the outlier data set with associated *p*-values (below the diagonal); bolded estimates are significant before correction, bolded estimates with asterisks are significant after correction.

	WL	SC	ML	DT	MG	PC	JC2	JC1	TX	VC	CH
WL	*	0.0123	**0.0315***	0.0137	0.0072	0.0115	**0.0261***	**0.0306***	**0.0460***	0.0144	**0.0143**
SC	0.0806	*	**0.0492***	**0.0226***	**0.0276***	**0.0196***	**0.0350***	**0.0224***	**0.0561***	0.0031	0.0047
ML	**0.0012***	**< 0.0001***	*	**0.0582***	**0.0504***	**0.0478***	**0.1037***	**0.0935***	**0.1236***	**0.0589***	**0.0553***
DT	0.0568	**0.0029***	**< 0.0001***	*	**0.0200***	**0.0244***	**0.0309***	**0.0325***	**0.0591***	**0.0186***	**0.0237***
MG	0.3700	**0.0023***	**0.0001***	**0.0260***	*	**0.0224***	**0.0451***	**0.0475***	**0.0539***	**0.0322***	0.0150
PC	0.0993	**0.0063***	**< 0.0001***	**0.0009***	**0.0114***	*	**0.0336***	**0.0401***	**0.0412***	**0.0168***	**0.0183***
JC2	**0.0005***	**< 0.0001***	**< 0.0001***	**0.0001***	**< 0.0001***	**< 0.0001***	*	**0.0202***	**0.0186**	**0.0328***	**0.0355***
JC1	**< 0.0001***	**0.0024***	**< 0.0001***	**< 0.0001***	**< 0.0001***	**< 0.0001***	**0.0038***	*	**0.0503***	**0.0306***	**0.0397***
TX	**< 0.0001***	**< 0.0001***	**< 0.0001***	**< 0.0001***	**< 0.0001***	**< 0.0001***	**0.04520**	**< 0.0001***	*	**0.0531***	**0.0351***
VC	0.0516	0.6213	**< 0.0001***	**0.0168***	**0.0015***	**0.0210***	**< 0.0001***	**0.0001***	**< 0.0001***	*	0.0063
CH	**0.0436**	0.4581	**< 0.0001***	**0.0014***	0.0811	**0.0086***	**< 0.0001***	**< 0.0001***	**0.0001***	0.3621	*

Presentations:

Puritz, J. B., Portnoy, D. S. and Gold, J. R. (2016) Mind the gap: the effects of INDELs and over-splitting on population genetic inference from RAD sequencing. Evolution, Austin, TX, June 2016.

Puritz, J. B., Portnoy, D. S. and Gold, J. R. (2016) Testing the genomic impacts of the DWH oil spill on red snapper Gulf of Mexico oil spill and ecosystem science conference, New Orleans, LA, February 2016.

Puritz, J. B., Portnoy, D. S. and Gold, J. R. (2015) Variable patterns of genomic diversity in young-of-the-year Red Snapper (*Lutjanus campechanus*) in the northern Gulf of Mexico. Annual Meeting of the American Fisheries Society, Portland, OR, November 2015.

Puritz, J. B., Hollenbeck, C. M. and Gold, J. R. (2015) Fishing for selection, but only catching bias: examining library effects in double-digest RAD data in non-model marine species. Plant and Animal Genomics, San Diego, CA. January 2015.

Puritz, J. B. and Gold, J. R. (2014) Genomic studies of red snapper (*Lutjanus campechanus*) in U.S. waters of the Gulf of Mexico and Atlantic Ocean. Evolution, Raleigh, NC 2014.

Puritz, J. B., Hollenbeck, C. M. and Gold, J. R. (2013) Genomic studies of red snapper (*Lutjanus campechanus*) in U.S. waters of the Gulf of Mexico and Atlantic Ocean. Annual Meeting of the American Fisheries Society Little Rock, AR, September 2013.

Publications:

Puritz, J.B., Hollenbeck, C.M, and Gold, J.R. (2014) *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ,* 10: e431.

Willis, S. C., Hollenbeck C. M., Puritz, J. B., Gold, J.R., Portnoy D.S. (2017) Haplotyping RAD loci: an efficient method to filter multi-copy loci and account for linkage disequilibrium. *Molecular Ecology Resources*, 17: 955-965.

Puritz, J. B., Gold, J. R., and Portnoy, D. S. (2016) Fine-scale partitioning of genomic variation among recruits in an exploited fishery: causes and consequences. *Nature: Scientific Reports*, 6, 36095. doi:10.1038/srep36095

Puritz, J.B., Gold, J.R. and Portnoy, D.S. (2017) The effects of over-splitting loci on population genetic inference from RAD sequencing. In preparation for *Molecular Ecology Resources.*

O'Leary, S.J., Puritz, J.B., Willis, S.C., Hollenbeck, C.M., Gold, J.R. and Portnoy D.S**.** (2017) These are not the loci you are looking for: Principle of effective SNP filtering for the molecular ecologist. In preparation for *Molecular Ecology*.