Population Genetic Analyses of Cobia within U.S. Coastal Waters

T. Darden, M. Walker, M. Jamison, M. Denson, W. Sinkus, K. Kanapeckas

SEDAR58-SID-04

Submitted: 3 April 2018
Updated: 16 April 2018

Please cite this document as:

Darden, T., M. Walker, M. Jamison, M. Denson, W. Sinkus, and K. Kanapeckas. 2018. Population Genetic Analyses within U.S. Coastal Waters. SEDAR58-SID-04. SEDAR, North Charleston, SC. 9 pp.

SEDAR58-SID-04

Population Genetic Analyses of Cobia within U.S. Coastal Waters

Darden, T., M. Walker, M. Jamison, M. Denson, W. Sinkus, and K. Kanapeckas
South Carolina Department of Natural Resources
Marine Resources Research Institute
April 3, 2018; updated April 16, 2018

Sample Collections:

The SCDNR Genetic Tissue Collection currently houses more than 5,050 archived genetic samples of cobia collected by numerous researchers and fishermen around the globe. In all cases, small tissue samples were collected from the pectoral, anal, or caudal fin and stored in either 95% non-denatured EtOH or a sarcosyl-urea preservation solution (8M urea, 1% sarcosyl, 20 mM sodium phosphate, 1 mM EDTA) until processing. For the current project, sample selection included those collected along the U.S. Gulf of Mexico and southeastern Atlantic coasts with available samples ranging from Virginia south along the Atlantic coast around the Florida peninsula into the Gulf of Mexico and westward to Texas. Three separate data sets were subjected to analyses: 1) All selected samples available, 2) Selected samples collected during cobia spawning season defined for each state on temperature-based patterns: Virginia – June to August, North Carolina – May to July, South Carolina and Georgia – April to July, Florida – March to August, Mississippi – May (only samples available), and Texas – April to August, and 3) Selected samples collected during cobia spawning season constrained to the period of April through June for all locations.

Genetic Protocols and Analyses:

The sarcosyl-urea preservative simultaneously stabilizes sample DNA and serves as a preliminary cell lysis solution. EtOH-stored samples were subjected to a proteinase K cell lysis overnight prior to DNA isolation. All DNA isolation, microsatellite amplification, and genotyping methods followed previous work on cobia from our group (Darden et al. 2014). Briefly**,** DNA was isolated from all samples using a metal bead isolation procedure. Ten polymorphic microsatellite loci were then amplified via polymerase chain reaction (PCR) in three multiplexed groupings. These loci have been optimized and multiplexed previously, and were used by our lab to document both global and local population structure in cobia. PCR was conducted in 11 µL reactions with 1x HotMaster buffer with 2.5 mM Mg^{2+}, 0.2 mM dNTPs, 0.3 units HotMaster *Taq* polymerase (5 Prime, Inc.), 0.5 mM $MgCl_2$, 0.20 mg/mL BSA, 0.3 µM forward and reverse primers, and 1 µL of 1:10 diluted DNA template. Individual primer concentrations differ among loci and are given in Darden et al. (2014). Forward primers for all loci were labelled with WellRED fluorescent dyes (Beckman Coulter, Inc.). Thermal cycling for PCR used a modified 60°C touchdown protocol (from Renshaw et al. 2006) consisting of an initial denaturation step at 94°C for 2 min, followed by 34 cycles of denaturing at 94°C for 30s, annealing at 60°C, 57°C, and 54°C (7, 7, and 20 cycles, respectively) for 1 min, and extension at 64°C for 2 min, followed by a final extension step at 64°C for 60 min (as in Darden et al. 2014). Both size standards (Genome Lab DNA Size standard kit 400) and reaction products were separated with a Beckman CEQ 8000 (Beckman Coulter, Inc.), with fragment size analysis performed with CEQ8000 software. All chromatograms were scored manually by two independent readers. Discrepancies between readers were resolved in conference, or samples were rerun to obtain an unambiguous genotype for all individuals.

As researchers at SCDNR initiated a cobia stock enhancement research program in 2004, the genetic samples collected for this project were also screened for hatchery individuals in the sampled populations. We utilized a maximum likelihood parentage approach as implemented in CERVUS 3.0.3 (Kalinowski et al. 2007) to provide a statistical evaluation of parentage taking into account mutation rates, population allele frequencies, and lab error rates. The power of the loci suite to correctly identify hatchery fish as well as individual fish is high, with average parent-pair and identity non-exclusion probabilities of $1.7x10^{-7}$ and $7.8x10^{-12}$, respectively, suggesting very low probabilities of incorrectly identifying hatchery fish or individuals. Parentage simulations (n=20) were run with known sex parentage analysis using allele frequencies from individuals collected from 2007 to 2009 (n=1,407). All simulations were conducted with 10,000 offspring, 8 candidate parent pairs (with all parents sampled), 95% genotyping, and low mistyping error (0.01) and mutation (0.001) rates. Critical delta scores were determined using 99% confidence for the relaxed criteria and 99.9% for the strict criteria. Parentage analyses for the juvenile samples were conducted with the modal simulation file from the simulation runs. All parental assignments were designated at the strict confidence level (99.9%). All hatchery-born fish were removed from the dataset prior to further analysis.

All remaining individuals that were successfully genotyped at 8 or more loci were subjected to sibship analyses as implemented in the software Colony 2.0.6.4 (Jones & Wang 2010) to identify any potential large family groups within the dataset that could confound further genetic structure analyses. Two simulations were run using settings of polygamous breeding, weak prior, updating allele frequencies, no genotyping error, and FPLS likelihood method for a medium run length. Any identified duplicate samples were removed from the dataset prior to further analyses. Results were evaluated for consistency among runs for individual fullsib relationships as well as family sizes present.

Standard population genetic statistical analyses were applied to the resulting sample datasets. Population genetic structure throughout the collection range was assessed via evaluations of Hardy Weinburg equilibrium (HWE) in GenAlEx 6.5 (Peakall & Smouse 2006, 2012), AMOVA analyses in Arlequin 3.5.1.2 (Excoffier and Lischer 2010), pairwise F_{ST}-style statistics calculated in GenAlEx 6.5 and Arlequin, and with the clustering algorithms implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000). Iterative AMOVA (R_{ST}-based) analyses were conducted to evaluate areas of genetic discontinuity in the data sets with potential location groupings under two-population scenarios. Pairwise comparisons of sample locations and HWE were conducted initially at the smallest geographic scale and locations were combined sequentially to represent the smallest number of homogenous groupings. Estimates of R_{ST}, F_{ST}, G_{ST}, G'_{ST} (Nei), G''_{ST}, and D_{EST} were initially calculated to verify consistency across metrics; as patterns of all estimates were consistent, only R_{ST} metrics are reported. The clustering model assignment employed in the program STRUCTURE using a hierarchical approach with the assistance of the web-based software Structure Harvester 0.6.94 (Earl et al. 2012) was used to identify the most appropriate number of distinct populations (K) of each run. Simulations were run with and without the locprior (collection location) parameter for all analyses, with five replicates for each K, the length of the burn-in period set at 20,000, and number of Markov chain Monte-Carlo reps after burn-in set at 20,000. All analyses were conducted from K=1 to K= # collection locations included +1. Sites that showed homogenous ancestry patterns were removed from the data set and STRUCTURE was run iteratively until K=1 was the most appropriate assignment for each cluster (based on combined evaluation of the Evanno method and log likelihood plots). Effective number of migrants per generation and year (based on 5-7 year generation time for cobia) were calculated for each resulting homogenous cluster in Arlequin.

Results and Conclusions:

Only a single duplicate sample and 39 cultured fish occurred within the dataset; these were removed from all datasets prior to analyses. No large family groups (>3) were present within the dataset and only 12 full sibling pairs were identified (p=1.0); therefore no confounding effects from family structure are anticipated in further analyses. A total of 2,954 samples were successfully genotyped for inclusion in the complete dataset, 2,796 samples met our by State selection criteria, and 2,508 samples were included in the April-June dataset (Table 1). Collection years for all samples included 2006 through 2017. For the initial analyses, the dataset was partitioned into 18 geographic sections based on natural latitudinal breaks in the collection data (Figure 1).

Table 1. Genetic dataset sample composition including location descriptions, locality codes, collection years of samples, and sample sizes for both the initial partitioned datasets and the final grouped datasets (see text for grouping methodology).

Code	Location definition	Collection Years	Sample Sizes			Final Sample Sizes		
			All Data	By State	Apr-Jun	All Data	By State	Apr-Jun
TX	Texas (Corpus Christi)	2010-2011	61	51	15	401	385	298
MS	Mississippi	2010	6	6	6			
FLW	FL Panhandle	2008, 2017	45	45	45			
FLS	Florida Keys	2010, 2015	9	9	9			
FLE1	Boynton Beach to Jupiter	2016-2017	39	36	26			
FLE2	Hobe Sound to Ft. Pierce	2011, 2015-2017	241	238	197			
FLE3	Canaveral/Sebastian (plus a few Ponce samples)	2014-2017	86	77	13	86	77	50
FLGA	Jacksonville, FL to Brunswick, GA	2009-2010, 2016-2017	16	16	5	52	50	
GA	Savannah	2008-2009, 2012, 2014-2016	36	34	32			
SCO1	Offshore Port Royal Sound and Betsy Ross reef	2009-2016	434	430	426	1412	1291	1193
SCO2	Offshore Charleston, Murrels Inlet, Georgetown	2007-2011, 2015-2017	39	21	18			
SCO	All other offshore samples from SC	2007-2009, 2014-2017	675	615	602			
NCO1	Offshore south of Cape Hatteras	2010, 2013-2014, 2016-2017	47	35	33			
NOC2	Offshore at and north of Cape Hatteras	2008-2010, 2016-2017	217	190	114			
SCI	SC inshore	2005, 2007-2016	835	834	831	835	834	831
NC1	Inshore area around Cape Lookout	2010, 2016-2017	20	16	16	168	159	136
NC2	Inshore area Pamlico Sound area	2010, 2016-2017	43	41	36			
VA	VA inshore	2006-2008, 2017	105	102	84			

Figure 1. Geographic distribution of the sample collection partitions. Size of the ellipses are not indicative of sample sizes at each location, but do indicate the range along the coastline from which fish were sampled.

All evaluated datasets resulted in consistent gene flow patterns following analyses, with the only deviation being a reduction in clarity (i.e., strength of patterns) within the North Carolina/Virginia collections in the April-June constrained dataset likely due to the loss of samples during the July-August time periods within this region that includes the known peak spawning period (Joseph et al. 1964). As such, we have selected the dataset with spawning season defined by state to present results within this working paper in order to minimize noise due to winter collected samples (high movement period) as well as maximize any potential signal from the North Carolina and Virginia areas. Additionally, all STRUCTURE analyses with and without the locprior parameter produced consistent results; therefore, only the results from the locprior-based analyses are presented here as the patterns are more clearly visualized in the STRUCTURE plots.

Multiple rounds of heirarchical STRUCTURE, initial pairwise F_{ST}, and HWE analyses supported a genetically distinct South Carolina inshore population (Figure 2) and a homogenous Gulf of Mexico population ranging from Texas through the Ft. Pierce, FL area (FLE2, Figure 3). Additionally, the Virginia and inshore North Carolina (NC1, NC2) samples represented a distinct genetic grouping (Figures 3 & 4) as did the combined offshore South Carolina and North Carolina samples (SCO, SCO1, SCO2, NCO1, NCO2, Figure 4). Samples from the Cape Canaveral, FL through Savannah, GA locations showed genetic similarities with collection locations from both the north (SCOs) and south (FLE2) and appeared to be a transition zone in the STRUCTURE analyses (Figure 3).

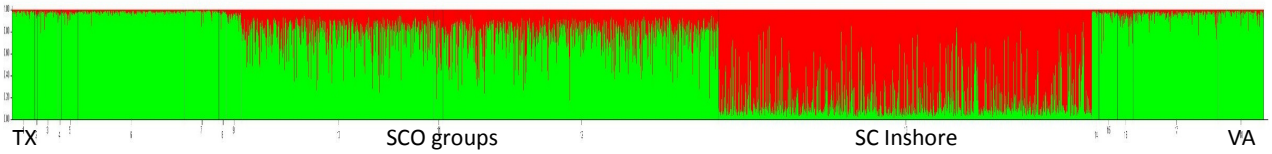

Figure 2. Population ancestry plot for the complete data set based on STRUCTURE results of K=2 with the South Carolina inshore collection identified as distinct from the remaining dataset (independent analysis confirmed K=1). Each vertical bar represents a single individual in the plot with colors indicating percent ancestry to each genetic group. Collections are geographically oriented from Texas on left to Virginia inshore on the right.

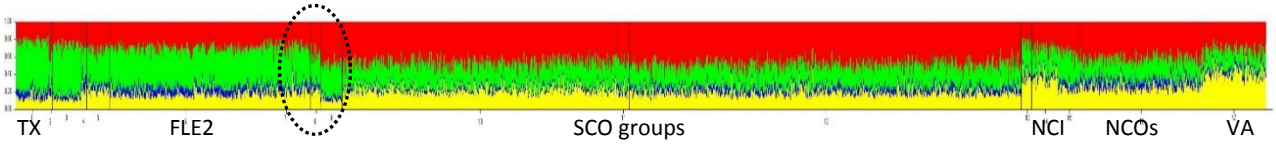

Figure 3. Population ancestry plot for the dataset excluding the South Carolina inshore samples. STRUCTURE results shown for K=4 with the Gulf of Mexico and Florida samples (through FLE2, Ft. Pierce) representing a distinct group (independent analysis confirmed K=1), North Carolina inshore and Virginia inshore samples representing a distinct group (independent analysis confirmed K=1), South Carolina and North Carolina offshore samples being homogenous (independent analysis confirmed K=1), and Cape Canaveral, FL through Savannah, GA representing a transition zone (dashed oval). Each vertical bar represents a single individual in the plot with colors indicating percent ancestry to each genetic group.

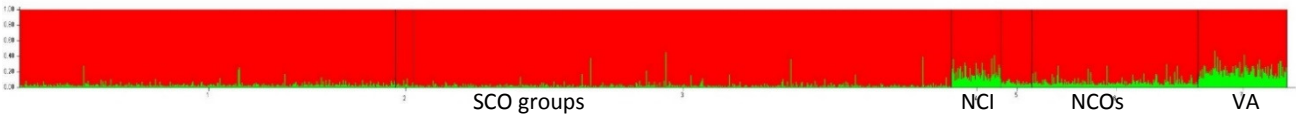

Figure 4. Population ancestry plot for the dataset excluding South Carolina inshore, Gulf of Mexico and east coast of Florida/Georgia. Collections are geographically organized with South Carolina offshore on the left and Virginia inshore on the right. STRUCTURE results shown for K=2 with the North Carolina inshore and Virginia inshore collections grouping together (independent analysis confirmed K=1) with the remaining South Carolina and North Carolina offshore collections being independently analyzed to confirm K=1 for an Atlantic offshore group. Each vertical bar represents a single individual in the plot with colors indicating percent ancestry to each genetic group.

As such, the iterative AMOVA analyses were employed to evaluate potential breaks in gene flow within the area including all potential locations from Atlantic offshore samples (SCOs, NCOs)/Savannah (GA) through Jupiter Beach (FLE1)/Hobe Sound (FLE2). Results suggested that the strongest significant break (R_{ST}=0.0073, p=0.001) among the groupings occurred between the Cape Canaveral, FL (FLE3) and Jacksonville, FL/Brunswick, GA (FLGA) locations, explaining 0.73% of the variation in the dataset. However, grouping scenarios between the Atlantic offshore (SCOs, NCOs) and Savannah (GA) samples, Savannah (GA) and Jacksonville, FL/Brunswick, GA (FLGA) samples, and Ft. Pierce, FL (FLE2) and Cape Canaveral, FL (FLE3) samples were also significant but not as strong (R_{ST}=0.0067-0.0069, p=0.004-0.006), explaining 0.67-0.69% of the variation. The last AMOVA scenario (break between Jupiter Beach and Hobe Sound) did not partition a significant amount of variation among groups (p=0.450). Therefore, the AMOVA analyses also support the occurrence of a transition zone from Cape Canaveral, FL through Savannah, GA.

Guided by these analyses, final sample groupings included Gulf of Mexico, South Carolina Inshore, North Carolina/Virginia Inshore, and Atlantic Offshore populations as well as the Cape Canaveral, FL and Jacksonville, FL-Savannah, GA groups (Table 1, Figure 5). Due to the lower samples sizes from some of the Gulf of Mexico and North Carolina inshore collection locations, deviation from HWE was evaluated to verify no substructure was being masked within these regions. No loci were out of HWE within the Gulf of Mexico group and only a single locus was out of HWE (p<0.001) in the combined North Carolina Inshore and Virginia dataset, supporting both of these groupings. Pairwise comparisons among these groupings confirmed significant differences between all groups (p<0.00001-0.04), except comparisons between the Cape Canaveral, FL and Jacksonville, FL/Savannah, GA groups with Gulf of Mexico and Atlantic Offshore populations (p=0.07-0.96). Significant genetic differentiation ranged from R_{ST}=0.020 between South Carolina Inshore and Gulf of Mexico populations to R_{ST}=0.006 between South Carolina Inshore and Atlantic Offshore populations (Table 2). The levels of genetic differentiation detected translated into effective number of migrants (N_{em}) ranging from 0.2-10 individuals per year between these populations (Table 2, Figures 6 & 7). Therefore, the results suggest the cobia stock boundary, recognizing that biologically this represents a transition zone with limited reproductive exchange, is occurring somewhere within the range from Cape Canaveral, FL to northern Georgia, which is consistent with the current management stock boundary along the Atlantic coast.

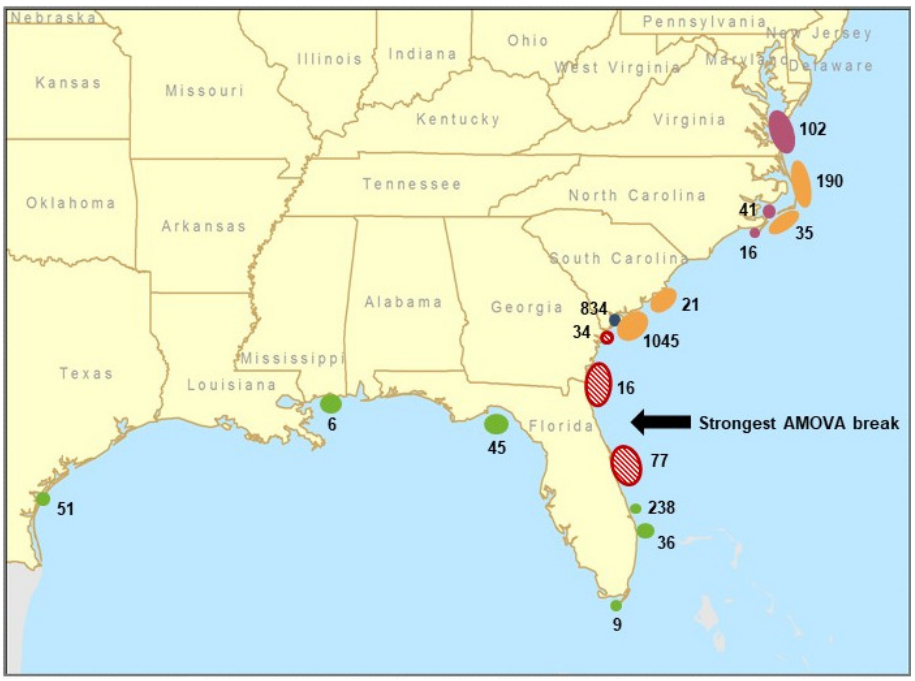

Figure 5. Visual summary of composite analyses of gene flow patterns for cobia along the U.S. Gulf of Mexico and southeastern Atlantic coasts. Solid colors represent populations which are genetically distinct from each other while the hatched, red locations remain unassigned to any genetic population. Sample sizes show are for the dataset with spawning season defined by state.

Table 2. Ranked R_{ST} values from significant pairwise comparisons among final genetic sample groups with effective number of migrants (N_{em}). Per year calculations are based on a 5-7 year generation time.

Pairwise Comparison	R_{ST}	N_{em} per generation	N_{em} per year
SC Inshore – GOM	0.020	1	0.2-0.6
SC Inshore – NCVA Inshore	0.019	12.9	1.8-2.6
SC Inshore – Cape Canaveral, FL	0.015	20.5	2.9-4.2
NCVA Inshore – GOM	0.012	20.6	2.9-4.1
NCVA Inshore – Jacksonville, FL/Savannah, GA	0.011	22.5	3.2-4.5
NCVA Inshore – Cape Canaveral, FL	0.009	27.5	3.9-5.5
SC Inshore - Jacksonville, FL/Savannah, GA	0.009	27.5	3.9-5.5
Atlantic Offshore – GOM	0.007	35.5	5.1-7.1
SC Inshore – Atlantic Offshore	0.006	41.4	5.9-8.3
NCVA Inshore – Atlantic Offshore	0.005	49.8	7.1-10.0

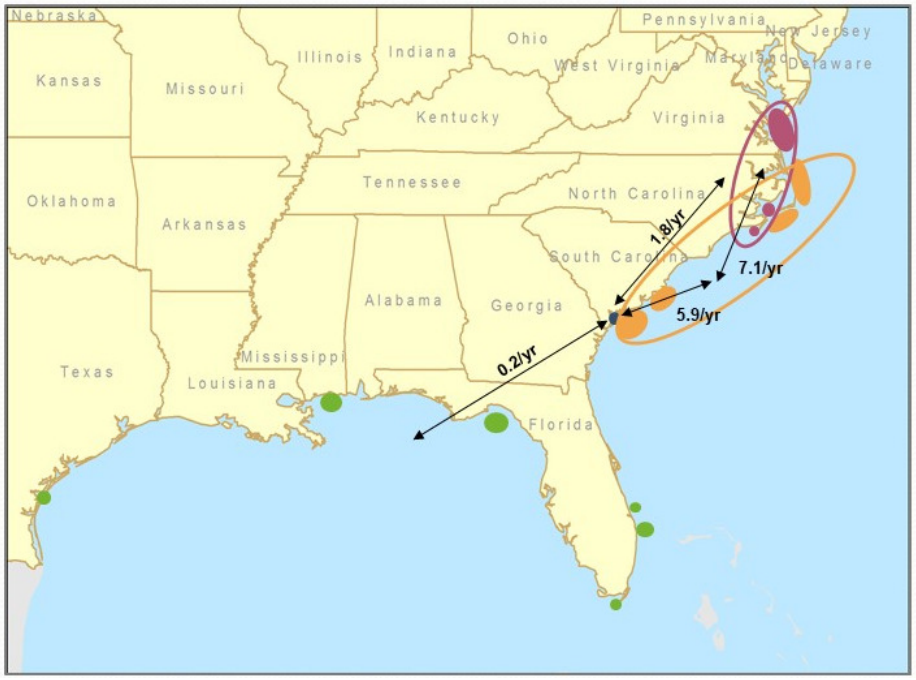

Figure 6. Visualization of the effective number of migrants per year (based on a 7 year generation time) between genetically distinct cobia populations along the U.S. southeastern Atlantic coast as well as between the South Carolina inshore and Gulf of Mexico populations.

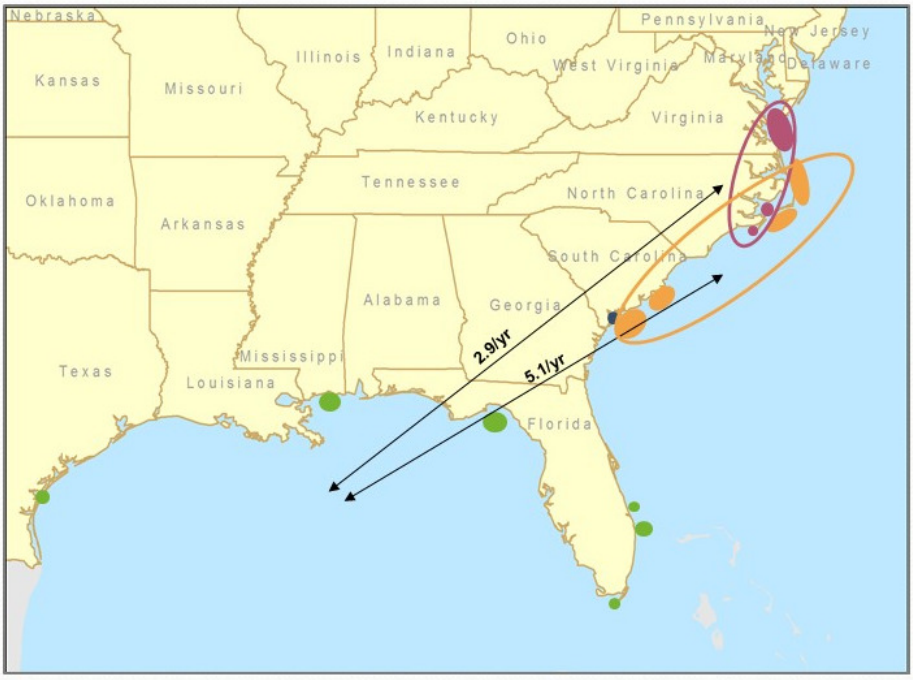

Figure 7. Visualization of the effective number of migrants per year (based on a 7 year generation time) between genetically distinct cobia populations along the U.S. southeastern Atlantic coast and the Gulf of Mexico population.

Literature Cited:

Darden, T.D. Cobia preliminary data analyses – US Atlantic and GOM genetic population structure. Working paper, SEDAR28-DW01. Submitted 5 January 2012.

Darden, T. D., M. J. Walker, K. Brenkert, J. R. Yost, and M. R. Denson. 2014. Population genetics of Cobia *Rachycentron canadum*: implications for fishery management along the coast of the southeastern United States. Fishery Bulletin 112:24–35.

Earl, D. A. and B. M. von Holdt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4(2):359-361.

Excoffier, L., and H.E.L. Lischer. 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.

Jones, O. and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10:551–555.

Joseph, E., J. Norcross, and W. Massmann. 1964. Spawning of the cobia, *Rachychentron canadum*, in the Chesapeake Bay area, with observations of juvenile specimens. Chesapeake Science 5(1-2):67-71.

Kalinowski S.T., M. L. Taper, and T.C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099-1106.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Peakall, R. and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539.

Pritchard J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Renshaw M.A., E. Saillant, S.C. Bradfield, and J.R. Gold. 2006. Microsatellite multiplex panels for genetic studies of three species of marine fishes: red drum (*Sciaenops ocellatus*), red snapper (*Lutjanus campechanus*), and cobia (*Rachycentron canadum*). Aquaculture 253:731-735.