

Genetic stock structure and connectivity of Queen Triggerfish (*Balistes vetula*)

Eric Saillant, John Horne, Luca Antoni

SEDAR80-RD-09

February 2022



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.

Final Report

**GENETIC STOCK STRUCTURE AND CONNECTIVITY OF
QUEEN TRIGGERFISH (BALISTES VETULA)**

Cooperative Research program award # NA14NMF4540062

Eric Saillant, John Horne, Luca Antoni

School of Ocean Science and Engineering, The University of Southern Mississippi, Gulf Coast
Research Laboratory, Ocean Springs, MS, 30564

November 29 2018

Executive summary

The queen triggerfish (*Balistes vetula*) supports commercial and recreational fisheries in the US Caribbean and is also a popular fish among recreational fishers and divers in the South Atlantic region. The perceived decline in abundance of queen triggerfish in many parts of its range and the intensive exploitation of this species in the US Caribbean have stimulated efforts to assess stocks to inform management of the fishery. The lack of essential data on life history, stock structure and demography throughout the species range in the US made attempts to assess this resource very difficult. Characterization of stock structure was identified as an immediate priority in order to develop appropriate data collection plans for future sound assessment of the resource.

This work addresses this need by providing a first assessment of genetic stock structure of queen triggerfish in US waters and one population in the Lesser Antilles. In a first project component, genomic resources were developed for queen triggerfish in order to improve inferences on stock structure, adaptive genetic variation and effective population size of regional populations during genetic studies. A draft reference genome sequence was generated using illumina sequencing. The obtained assembly spans over 778 Mb including 344 Mb in contigs over 1,000 bp and is expected to encompass most of the queen triggerfish genome which was estimated to be 575 Mb in length. The assembly is fragmented with a N50 of 1,515 bp and a L50 count of 100,353. Genome contigs were anchored during a linkage mapping experiment where 200 queen triggerfish larvae from a spontaneous spawn were genotyped using the double-digest Restriction Associated DNA (dd-RAD) sequencing method. A total of 15,027 loci could be ordered on the map yielding an average interval between adjacent markers of 0.09 cM. The map anchors 10,367 genome contigs spanning a total length of 19,95 Mb.

Genetic variation among queen triggerfish from 4 localities in the US Caribbean (East and West coasts of Puerto Rico, St Croix and St Thomas), one locality on the Southeast Atlantic coast (Jupiter, FL sampled in 2012 and 2014) and the French Antilles (La Martinique) was assessed (n = 67-89 per sample). Specimens were obtained in collaboration with commercial fishermen and charter boat captains in all localities. Samples were assayed using the double digest Restriction Associated DNA sequencing protocols and sequencing reads mapped to the draft reference genome assembly for analysis. The obtained dataset was filtered according to two

strategies, the first one aiming to maximize the density of the genome scan (yielding 4,512 loci) and the second to maintain large sample sizes per locality (but yielding only 1,698 loci). Outlier analyses conducted using two approaches did not provide consistent support for divergent selection affecting any of the surveyed loci. Genetic variation among populations was weak with all pairwise F_{ST} estimates lower than 0.0083. Parameters estimated from the isolation by distance model indicated dispersal over long distances was occurring (point estimates of the mean dispersal distance between 581 and 1,773 km) consistent with the extended duration of pelagic dispersal reported for this species (over 80 days). Effective population size estimates differed among localities suggesting different demographic assemblages contributed to recruitment in individual regions. Estimates were highest for localities located on the northern Puerto Rican platform (East and West of Puerto Rico and St Thomas) where the species is most abundant and heavily exploited, and lowest in La Martinique, Saint Croix and South Florida. The two samples from South Florida differed in allele frequencies and in estimates of effective population size suggesting recruitment in that region may be under the influence of different spawning assemblages contributing juveniles through pelagic transport. Analysis of recent demographic trajectories using the linkage disequilibrium method suggested effective population size had declined in all surveyed geographic locations and may be insufficient to ensure long-term sustainability of regional populations. The signal of decline tended to be more pronounced in La Martinique and Jupiter consistent with the lower abundance of the species in these two areas.

1. Purpose

The main purpose of this project was to document stock structure, connectivity and genetic demography of the queen triggerfish in US waters. However, considering the high potential for dispersal of this species that features an extended pelagic larval phase, the connectivity of US stocks with other populations in the Caribbean was also considered during evaluation.

1.1 Stock structure analysis

Knowledge of stock structure and connectivity is central to fisheries management. Failure to recognize distinct sub-units of an exploited stock may lead to overexploitation and depletion of local units (Carvalho and Hauser 1994, Begg et al. 1999), lack of success of management (Stephenson 1999), and/or inability to anticipate and measure recruitment success in fisheries during rebuilding programs (Ruzzante et al. 1999). Local depletion and/or extinction also leads to the loss of unique genetic characteristics of regional units such as disease resistance, reproductive traits, tolerance to specific environmental conditions, or other traits involved in local adaptation. The loss of these characteristics affects the sustainability of local populations but ultimately also that of the overall metapopulation (Waples 1995). For these reasons, genetic stock structure analysis should focus on identification of genetically distinct stocks (Hilborn et al. 2003) and include assessment of local genetic adaptation. The latter is usually challenging because phenotypic differences among populations may be due to plasticity rather than adaptation and also because the traits affected by local adaptation are very often unknown *a priori*. The analysis of stock structure including the detection of local genetic adaptation is however greatly enhanced by the application of high-density genome scans obtained through next generation sequencing as these scans allow detecting outlier genomic regions affected by divergent selection and, once these are identified, the analysis of neutral and non-neutral processes. Until recently, the deployment of such genome scans was cost prohibitive in non-model species. However, recent developments of genotyping-by-sequencing protocols based on next-generation sequencing technologies (Baird et al. 2008, Glaubitz et al. 2014) now allow implementation of high resolution genome scans for reasonable costs (Allendorf et al. 2010, Avise 2010, Angeloni et al. 2012). The interpretation of the results of genome scans in non-

model species such as the queen triggerfish is greatly enhanced when information on the degree of physical linkage between markers is available. Development of a reference genome sequence and a linkage map, and mapping SNP loci on this reference prior to interpretation of spatial patterns of genetic variation therefore greatly improves the reliability of inferences on population structure. This approach enables identification of clusters of linked loci departing from expected patterns of divergence under the neutral distribution (outlier loci) signaling regions of the genome possibly undergoing natural selection. This approach also allows controlling for potential confounding effects of selection and physical linkage during inferences on the number and location of genomic regions involved in local adaptation of populations. The analysis of multi-locus patterns of variation (see e.g. Pritchard et al. 2010) and inferences on demographic parameters are also improved by reducing and controlling the bias resulting from unequal contributions of genomic regions to the dataset. Accounting for linkage between loci also allows inferring recent effective population size trajectories (Larson et al. 2014, Barbato et al. 2015, Hollenbeck et al. 2016).

1.2 Queen triggerfish distribution and fisheries

The queen triggerfish is a member of the Balistidae family found on tropical and subtropical reef habitats of the Atlantic basin. Although the species has been described in the Eastern Atlantic from Cape Verde and the Azores (Harmelin-Vivien and Quérou 1990) to Angola (Smith and Heemstra 1986), recent reports of the species in the eastern Atlantic are sparse (K. Michalsen, Institute of Marine Research, personal communication) and mostly consist of country records that are not accompanied by voucher specimens or detailed information on criteria used to confirm species identification diagnoses. In the western Atlantic, queen triggerfish are most abundant in the Caribbean Sea, off the Southeast coast of Florida, and off the central coast of Brazil. However, the abundance of queen triggerfish in the latter region is currently very low (C. Albuquerque, Universidade Federal do Espírito Santo, Brazil Personal communication). The species is also very infrequent in US and Mexican waters of the Gulf of Mexico (National Marine Fisheries Service, Fisheries Statistics division, A. Aguilar-Perera, Universidad Autónoma de Yucatán personal communication) (Figure 1) and in Central America (Belize: M. Gongora, Belize Fisheries Department, Personal Communication; Colombia: species listed as endangered,

Mejía and Acero 2002; Panama: C. Vergara, Universidad Tecnológica de Panamá Personal communication; Venezuela: F. Arocha, Universidad de Oriente Personal communication). Along the East coast of North America, although the species has been reported as far North as Canada (Scott and Scott 1988) or Massachusetts (Robins and Ray 1986), catches are anecdotal North of Florida (National Marine Fisheries Service, Fisheries Statistics Division, Personal Communication). Accordingly, the geographic area where queen triggerfish can be found in abundance compatible with sampling for population genetic studies appears to be restricted to the Antilles and the southern tip of Florida.

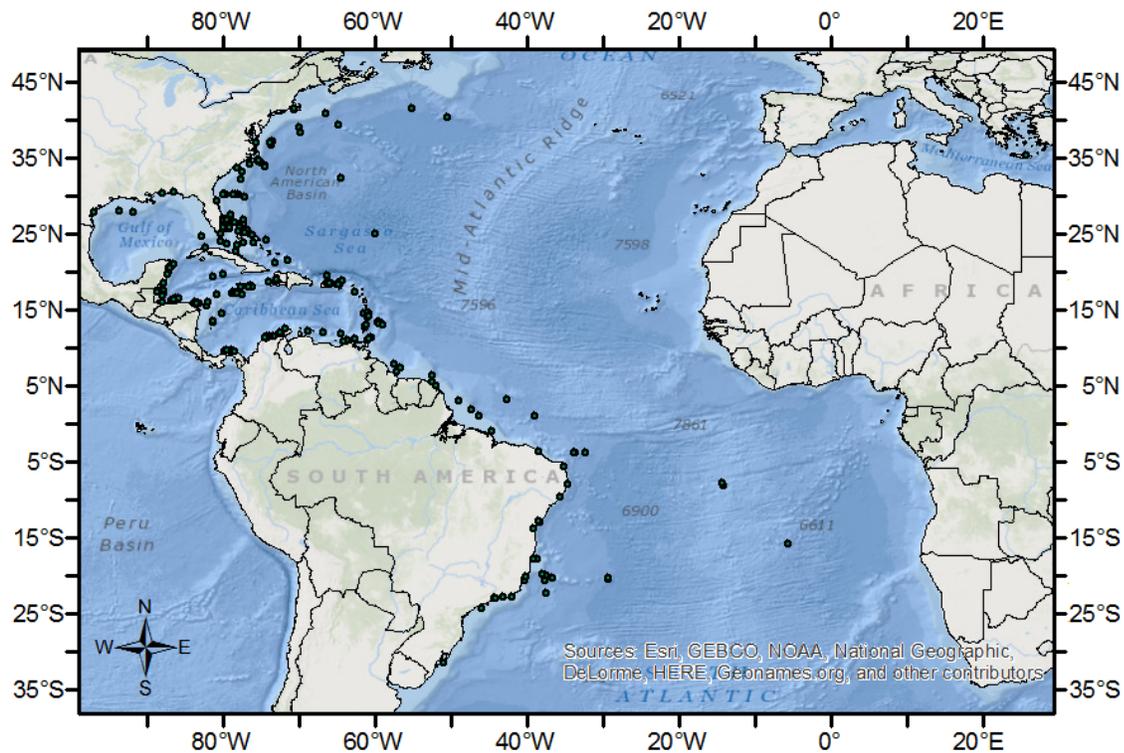


Figure 1 Queen triggerfish distribution map

Queen triggerfish (*Balistes vetula*) distribution range based on reviewed point observations. www.aquamaps.org, version of Aug. 2010. Updated 02/21/13.

Commercial landings of the species in the continental US are moderate ranging between 200 and 7,885 lbs between 2002 and 2016 (National Marine Fisheries Service, Fisheries Statistic Division, Personal communication). Major commercial fisheries occur in the US Caribbean where harvests have increased during the past few decades in part owing to the decline in abundance of trunkfish (Rosario et al. 1986). Queen triggerfish is the predominant species in St Thomas commercial fisheries (St. Thomas Fishermen’s Association, personal communication)

and is regularly ranked in the top 5 and top 10 of Sainte Croix and Puerto Rico for commercial finfish landings respectively. The species is traditionally caught using traps by commercial fishermen but, due to the reduction of the use of this gear in Puerto Rico, hook and line and spear-fishing are becoming the main capture methods for both commercial and recreational fishers. Queen triggerfish are sold on local food markets and are also traded on the aquarium market (Aida Rosario, personal communication).

Statistics on recreational catches reveal high fluctuations in landings over time. These fluctuations could be due in part to inconsistent data reporting, in particular for the Caribbean region, but overall available data suggest a clear and alarming declining trend since the 1980s (Figure 2, National Marine Fisheries Service, Fisheries Statistics Division, Personal communication).

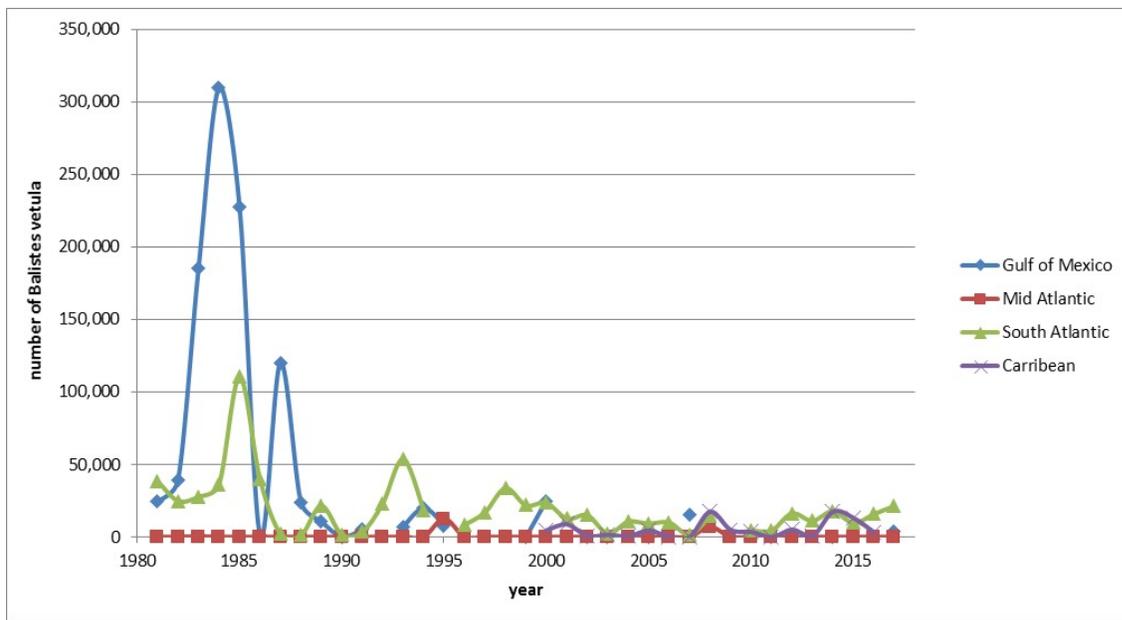


Figure 2. Recreational catch statistics for queen triggerfish in US waters by region for the period 1981-2017 (Source National Marine Fisheries Service, Fisheries Statistics Division, Marine Recreational Information Program).

Concerns regarding the status of queen triggerfish are not limited to US populations as this species was classified as Vulnerable by the International Union for the Conservation of Nature

(Roberts 1996) due to an estimated major decline in abundance over the past couple of decades. More recently, the species status was updated to the category “Near Threatened” on the basis of new estimates of populations decline over the distribution range which were lower than the previous ones and did not meet the 30% reduction in abundance threshold used to list species in the Vulnerable category (Liu et al. 2016). The main identified threat to the species is its level of exploitation (Roberts 1996) and the need for further evaluation is noted. Queen triggerfish is indeed classified as endangered in Columbia (Mejía and Acero 2002) and shows major declines in other parts of the Caribbean which could be a factor explaining the decline in abundance in the South Atlantic region if recruitment depends in part on transport of pelagic larvae from Caribbean geographic stocks.

Management of Caribbean reef fisheries has been of concern during the past few decades due to perceived declines of abundance for several species (Cummings and Matos-Caraballo 2004). Efforts to manage fisheries are limited by the insufficiency of data for individual species such as the queen triggerfish. A first assessment of queen triggerfish was conducted in 2012 (Rios 2012, SEDAR 30) and the panel was very limited in their effort due to the lack of essential data on the species particularly on demography and life history. Both the SEDAR 30 report and subsequent reviews concluded that sound and robust assessment could not be achieved with existing data and that better information on demography, life history, and biology of the species were needed. The panel emphasized the high priority need to develop data collection plans that would allow reliable assessment of this resource in the future. Specifically, the complete lack of information on stock structure was clearly apparent. Knowledge of stock structure and connectivity among US stocks and between US stocks and other Caribbean stocks is however essential in order to identify management units which is a pre-requisite to develop appropriate data collection plans to assess and manage these units.

1.3 Inferences from population genetic surveys

A first objective of the analysis of spatial genetic variation is usually to identify demographically independent stocks. Marine populations are often connected over large geographic distances and display an isolation-by-distance pattern of population structure (Rousset 1997), where genetic

distance increases as a function of geographic distance owing to limited dispersal potential (e.g. due to a limited duration of the pelagic larval transport in several reef fishes). While interpretation in terms of geographic boundaries of stocks are irrelevant under isolation by distance, indirect measures of dispersal distance (neighborhood size parameter that represents the dispersal distance an individual is assumed to travel from its birthplace to spread its genes, Rohlf and Schnell 1978, Fox et al. 2001) can be derived from the parameters of the isolation by distance model allowing to infer the degree of demographic dependence of geographic stocks of interest. However, variation among population is not always explained entirely by an isolation by distance model of population structure. When this is the case, deployment of Bayesian clustering approaches (e.g. Pritchard et al. 2000) or spatially explicit models allow detecting genetic discontinuities and delineating demographic units separated by barriers to gene flow (e.g. Dupanloup et al. 2002, Manni et al. 2004); when there is isolation by distance, this is best achieved accounting for spatial autocorrelation of samples within units as implemented in the approach developed by Francois et al. (2006) and Durand et al. (2009).

In species forming connected metapopulations as is often the case for marine organisms, information on demographic independence can also be gathered from the analysis of demographic parameters such as the effective population size which represents the size of an ideal population that experiences the same amount of genetic change as the population under study and is the appropriate population genetics parameter for characterizing available genetic resources of stocks in terms of evolutionary potential and long term viability (Hare et al. 2011). Independence even in the presence of gene flow and homogeneous allele frequency distributions can be inferred when contemporaneous estimates of N_e differ among regional populations. Information on the degree of independence among stocks can also be obtained from the analysis of the distribution of relatedness coefficients of individuals and the observation of a higher degree of relatedness of individuals within a deme than those sampled in different demes (Horne et al. 2016). The latter approaches are particularly helpful when dealing with highly connected species such as the queen triggerfish.

Another important inference in stock structure analysis is the detection of local adaptation of regional population if it occurs. This can be achieved through outlier analyses (Whitlock and Lotterhos 2015, Luu et al. 2017) and analysis of patterns of divergence at putatively selected

loci. Inferences are greatly improved when candidate outlier loci can be positioned in the genome in a sliding window analysis (e.g. Bourret et al. 2013, Holenhe et al. 2010).

1.4 Life history of queen triggerfish and potential for dispersal

Adult queen triggerfish are found on a broad spectrum of coral reef habitats ranging from shallow grassy habitats to the upper slope of deep reefs (Randall 1967) with depths ranging from 3 to 300 m although most specimens are found in waters less than 30 m deep (Gasparini and Floeter 2001). Age and growth in the Caribbean was documented by Manooch and Drennon (1987) with a maximum recorded age of 7 years. The minimum reported size at maturity is 17.5 cm (Fork Length) for males and 16.5 cm for females (Aiken 1983). The latter life history parameters are currently being updated (Riviera et al. 2014). In the Caribbean, queen triggerfish with ripe gonads are found primarily from February to May with a second period of reproductive activity reported from August to October; fish are usually not found in reproductive condition from October to January (Aiken 1983). Queen triggerfish are thought to spawn on the bottom, likely in a nest, as do most other reef associated members of the family Balistidae (Aiken 1983). They form spawning aggregations but are not assumed to conduct extensive migration to reach these aggregations (Richard Nemeth, Personal communication). Larvae are pelagic for a long period (settlement was reported to occur at 63-83 days old by Robertson 1988, corrected by Lindenman et al. 2000). The early life history of queen triggerfish thus indicates a high dispersal potential during the planktonic transport phase. Once settled, tagging studies indicate limited movement (Aiken 1983) although the size distributions observed in catches at different depths suggest that larger fish tend to move to deeper waters. The sedentary lifestyle of adult queen triggerfish makes the species potentially vulnerable to overfishing and significant modification of the age and size structure of local populations may be induced by fishing activities. Inference on stock structure based on life history or phenotypic traits (e.g. growth rate) are therefore likely unsuited, as was suggested for the closely related gray triggerfish, because growth characteristics of local populations may reflect primarily the degree of fishing pressure in that population (Ingram 2001).

In summary, the potential for dispersal of queen triggerfish is high during early life because of the long duration of the pelagic phase where larvae can be transported by oceanic currents. However, cyclonic or anti-cyclonic currents may retain most young triggerfish in the area where they were spawned leading to mostly self-recruited local populations as proposed for the gray triggerfish (Ingram 2001, SEDAR #9). A rigorous examination of genetic stock structure appeared therefore warranted with interpretations conducted accounting for predictions on the pelagic dispersal of larvae by surface currents.

1.5 Specific hypotheses in the context of regional management of the Caribbean and South Atlantic fisheries and project objectives

Examination of the patterns of prevailing surface currents, available from the Cooperative Institute for Marine and Atmospheric Studies (CIMAS) at <http://oceancurrents.rsmas.miami.edu/mgsva/slideshow.html?imgnum=6&basin=atlantic>, led to the formulation of the following hypotheses on the connectivity of geographic populations of queen triggerfish in the Caribbean region:

- a. The Antilles current flowing from the Lesser Antilles to the Bahamas is hypothesized to connect island populations located in its path. Dispersal of larvae through pelagic transport along this current is predicted to lead to a unidirectional (Southeast to Northwest) pattern of migration and isolation-by-distance along this axis. This current is very slow (4-9 cm/s) and discontinuous (Lee et al. 1996), which is expected to result in moderate dispersal distances (average dispersal would range from 218 km to 645 km considering a 63-83 days duration of the pelagic phase respectively) and increased demographic independence and divergence among geographically distant populations.
- b. The Bahamas and Florida peninsula are separated by unsuitable adult habitat (water depth reaches 6,000 ft) unlikely to be crossed by adult queen triggerfish. The fast-flowing Florida current runs through this trench, in an orientation parallel to the coastline and would therefore not promote exchange of larvae between the two stocks (Figure 3a), leading to a predicted barrier to migration between the Bahamas and the Florida East coast.

c. The Caribbean current and the Loop current are hypothesized to connect the northern countries of South America (e.g. Venezuela and Colombia) to Belize and South Florida although the Panama-Colombian Gyre may contribute to isolate populations in that region from those in the rest of the Caribbean (Figure 3b) as suggested for other reef species (Cowen et al. 2006). As in the case of the Antilles current, a unidirectional (South to North) pattern along the current axis is expected.

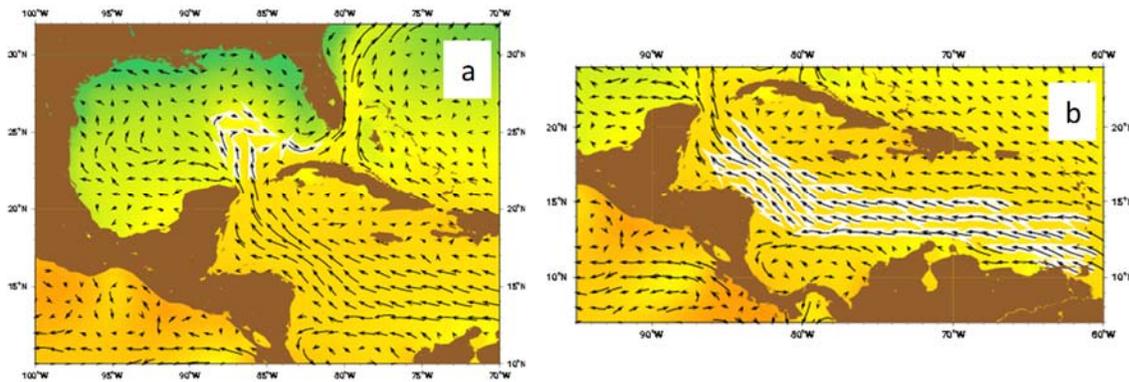


Figure 3. Schematics of prevailing currents in the Northern Caribbean, Gulf of Mexico and South East Florida (a) and the Southern Caribbean (b). Data correspond to averages during the period January to March corresponding to the peak spawning season of queen triggerfish (source The Cooperative Institute for Marine and Atmospheric Studies)

Additional hypotheses on the fine-scale stock structure among regional populations of queen triggerfish in the US Caribbean and South Florida can also be formulated. The South Atlantic region has been treated as an independent stock in assessments of Lutjanids with pelagic larvae (e.g. the yellowtail snapper, SEDAR-8 2005) considering the large geographic distance between South Florida and the US Caribbean which was assumed to be incompatible with direct larval transport between the two regions and the possible barrier between the Bahamas and East Florida discussed above (hypothesis b above).

In the US Caribbean, two stocks have been distinguished in reef fisheries assessment efforts. The first Caribbean stock corresponded to the northern Puerto Rican Platform that includes the islands of Puerto Rico, St. Thomas, St. John, and the British Virgin Islands, while the second

stock corresponded to the southern St. Croix Platform that includes the island of St. Croix. These two geologic platforms are separated by a deep trench that extends to ~4,000 meters (CFMC 2004) likely preventing adult reef fish (including queen triggerfish) migration. Considering a two-month period of larval drift (a lower bound for queen triggerfish larvae as discussed above), areas off of Puerto Rico and the US Virgin Islands could be highly connected via larval transport. There are however a number of caveats to this prediction. Cowen et al. (2000, 2006) using advanced biophysical models showed that effective larval dispersal in a variety of reef fishes was restricted to distance in the order of 10-100 km. In addition, the direction of dominant currents may lead to unidirectional gene flow (e.g. East to West in the US Caribbean) and promote local larval retention on the leeward side of islands while windward sides would be more prone to receiving migrants (Swearer et al. 1999).

Evidence of partially restricted gene flow between populations in the region was indeed recently documented in genetic studies conducted in three reef-associated Lutjanids (Carson et al. 2011; Gold et al. 2011; Saillant et al. 2012). These studies were consistent with demographic independence of the Caribbean and South Atlantic populations and also with a possible demographic independence between St. Croix and the northern platform, and between the eastern (East Coast of Puerto Rico, St. Thomas) and western sides of the northern platform (West coast of Puerto Rico).

Because of the unavailability of samples of queen triggerfish in the Southern Caribbean and South America, due to low and declining abundance of the species, this project focused on the axis Antilles – South Florida, where multiple locations could be sampled. This study examined the hypothesis that populations are structured according to an isolation-by-distance model, and evaluated the potential occurrence of finer patterns of population structure involving demographic independence of populations occupying the two Puerto Rican platforms or partial isolation of populations located on the leeward side of the US Caribbean.

The potential for local adaptation remains undocumented in queen triggerfish and was examined in this work through analysis of genome wide variation.

A last question of interest regarded the demographic dynamics of US regional stocks. Catches in the US Caribbean showed an increasing trend in the numbers of larger individuals caught in recent years, particularly along the East coast of Puerto Rico, but this trend may also reflect increased fishing effort. Indirect measures of abundance and changes in population size based on effective population size and its recent trajectory were therefore examined in this study.

The project proposal initially envisaged using microsatellite markers for the study but the project team decided to use the double digest Restriction Associated DNA (dd-RAD) sequencing protocol (Peterson et al. 2012) instead in order to achieve a high-density scan of the queen triggerfish genome as introduced in section 1.1. Because most of the inferences that can be made from genome scans are facilitated by the use of a reference genome, as also discussed in section 1.1, part of this work focused on developing genomic resources for queen triggerfish. A draft reference genome assembly based on short Illumina sequencing reads was developed and a linkage map was produced through analysis of the segregation of loci in progeny from a cross produced at the New England aquarium. This report is therefore organized in two parts addressing two objectives. In a first part, the development of genomic resources for queen triggerfish is described and a second part focuses on the survey of genetic variation and stock structure analysis in queen triggerfish populations of the Antilles and South Florida.

2. Approach

2.1 Development of genomic resources

2.1.1 Genome sequencing and assembly

Genome sequencing was conducted using the DNA from one queen triggerfish specimen collected offshore Jupiter, Fl. DNA was isolated using the qiagen DNA-easy kit® and submitted to the University of Colorado-Denver genomics and microarray core for sequencing on one full lane of the Illumina Hi-seq 2500 platform.

Obtained reads were quality filtered and trimmed with trimmomatic v. 0.36 (Bolger et al. 2014) and assembled with the SparseAssembler (Ye et al. 2011).

The size of the queen triggerfish genome was estimated from the quality filtered reads using k-mer parallel counting in Jellyfish (Marcais and Kingsford 2011).

2.1.2 Mapping family production and DNA isolation

Linkage mapping was conducted using progeny from a putative full sib family produced at the New England Aquarium (Boston, MA). Embryos were collected following a spawning event and incubated until hatch by personnel at the New England aquarium. Newly hatched larvae were fixed in ethanol 95% and shipped to USM for analysis. DNA was isolated from larvae using the omega Mag Bind kit®. Because the DNA yield was insufficient for the dd-RAD library preparation protocol, the genome of all larvae was amplified using the Qiagen Repli-g mini amplification kit and 200 larvae yielding 1.5 µg or more of DNA were selected for dd-RAD sequencing.

2.1.3 Double-digest Restriction Associated DNA sequencing

DNA concentration for each individual sample was determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and the quality of the DNA was also evaluated during agarose electrophoresis. DNA solutions were concentrated or diluted in order to achieve a final concentration of 40-100 ng/µL. dd-RAD sequencing libraries were prepared from each individual DNA using protocols modified from Baird et al. (2008) and Peterson et al. (2012). Each library was made from 0.52-1.3 µg of genomic DNA that was simultaneously digested with the restriction enzymes Sau3AI (3.75 units, New England Biolabs Inc. [NEB], Ipswich, MA, USA) and SPEI (7.5 units, NEB) at 37°C for 1.5 h. Enzymes were then heat-inactivated at 65°C for 15 min and the digested fragments were ligated at 16°C for 30 min to sample-specific Illumina adapters using 400 units of T4 Ligase (NEB). In order to identify sequencing reads originating from a specific individual sample, a unique 6 bp barcode was included in each adapter pair. Adapters also included a degenerate 8 bp unique molecular identifier (UMI) which was used to identify and remove PCR duplicates during processing of

sequencing reads as described by Schweyen et al. (2014). Ligation reactions were terminated by diluting 5 folds with ddH₂O. The products of the ligation reaction were then purified and size-selected to retain fragments larger than 250 bp using Agencourt AMPure XP magnetic beads (Beckman Coulter Inc., Brea, CA, USA). Beads were applied at a 0.65x ratio and unbound DNA was removed during two washes with 70% ethanol. Cleaned DNA products (2 μ L in volume) were then amplified through PCR in 25 μ L reactions consisting of 5 pmol of forward and reverse primers and 12.5 μ L of 2X Master Mix, containing Taq polymerase (50 units/mL), dNTPs (0.4 mM), MgCl₂ (3 mM), and 2X buffer (NEB). The thermal cycle consisted of an initial denaturation step of 3 min at 95°C, 30 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, and a final extension at 72°C for 7 min. PCR products from individual libraries were pooled for multiplex sequencing (75-100 individuals were pooled in each Illumina Hi-seq sequencing lane). The pool of libraries was cleaned with Agencourt AMPure XP beads as described above and libraries were size-selected on a PippinPrep® (2% agarose gel cartridge, Sage Science, Beverly, MA, USA) to isolate fragments ranging in size between 300 and 500 bp. The products of the elution step were purified using a 1.8x ratio of Agencourt AMPure XP beads and submitted to the University of Colorado, Denver School of Medicine (Aurora, CO, USA) for 2x125 paired-end sequencing on the Illumina HiSeq2000 platform.

2.1.4 Construction of a high density linkage map

Demultiplexed reads were screened for PCR duplicates. Duplicates were inferred when multiple reads showed identical sequences and shared the same UMI. Because PCR duplicates lead to an increase in homozygosity and an inflated confidence in genotype calls (Andrews et al. 2014), only one read from each set of detected duplicates was kept for further analysis. A quality check was then performed on the Illumina raw reads to reduce the probability of base-calling errors using a sliding window approach in Fastp (Chen et al. 2018). Paired-end reads were trimmed if the mean Phred-score in a 15 bp window fell below 20. The resulting sequences were discarded from the dataset if either the forward or the reverse read was shorter than 15 bp.

Filtered reads were processed using the dDocent pipeline (Puritz et al. 2014) for mapping on the reference genome in BWA-MEM (Li 2013). Sixteen individuals were removed from the dataset at this stage due to low coverage (< 1.5 M reads). The software FREEBAYES (Garrison and Marth

2012) was used to discover and genotype Single Nucleotide Polymorphisms (SNPs) and other variants such as insertions/deletions (INDELs), multi-nucleotide polymorphisms (MNPs), and complex events (composite insertions and substitutions). INDELs and complex events were excluded and only remaining loci with quality scores exceeding 30 were retained.

Linkage mapping was conducted in Lep-MAP 3 (Rastas 2017). Markers with less than 90 missing data and showing no significant segregation distortion were assigned to linkage groups using a threshold LOD score of 15. The obtained linkage groups were ordered 10 times and markers creating long gaps at the end of a linkage group (10 cM or higher) were removed to generate the final map

2.2 Population genomic survey

2.2.1 Sampling

Sampling aimed to assess population structure in the US Caribbean (Puerto Rico and US Virgin Islands) where the species is particularly abundant and can easily be sampled. Samples from four locations were obtained in this area in order to assess genetic variation in the region: East and West coasts of Puerto Rico and St. Thomas Island on the northern Puerto Rican platform and St. Croix Island on the southern platform. Sampling also aimed to describe genetic variation along the Antilles axis (Lesser Antilles to Greater Antilles) and evaluate a possible isolation-by-distance pattern. Samples from the Lesser Antilles were pursued and obtained from La Martinique Island. Samples from further South (Trinidad, Venezuela, Panama, and Colombia) were pursued without success due to the rarity of queen triggerfish in these waters. Samples were also pursued from Bahamas without success. Two samples were obtained from the Southeast of Florida (Jupiter) and included in the study to examine temporal stability at this location and confirm results initially obtained with the 2014 sample.

A total of 545 specimens were collected between 2012 and 2015 (Figure 4). Sampling focused on the winter peak of the spawning season to describe the breeding structure and involved commercial fishermen and charter captains. Individuals from Jupiter, Florida (JP, n = 80 in 2012 (JP1) and 79 in 2014 (JP2)) were collected by Capt. W. Taylor via hook and line. Specimens

from East Puerto Rico, near Fajardo (FA, n = 79), were obtained with the help of Capt. M. Hanke and caught via hook and line and using traps. Samples from West Puerto Rico, off Mayaguez (MAY, n = 75), were caught via spearfishing by Capt. Benigno Rodriguez in collaboration with the Fisheries Research Laboratory of the Department of Natural Resources. Individuals from Saint Thomas (ST, n = 87) were collected by Capt. Daryl Bryan. Saint Croix (SCR, n = 79) specimens were collected by trapping in collaboration with Capt. Thomas Daley. Finally, individuals from La Martinique Island (MA, n = 66) were provided by the institute IFREMER. Fin clips were collected from fresh fish at the dock or directly on the boat and preserved in a DMSO salt-saturated fixative (0.5 M EDTA, 20% Dimethyl sulfoxide, NaCl, ddH₂O).

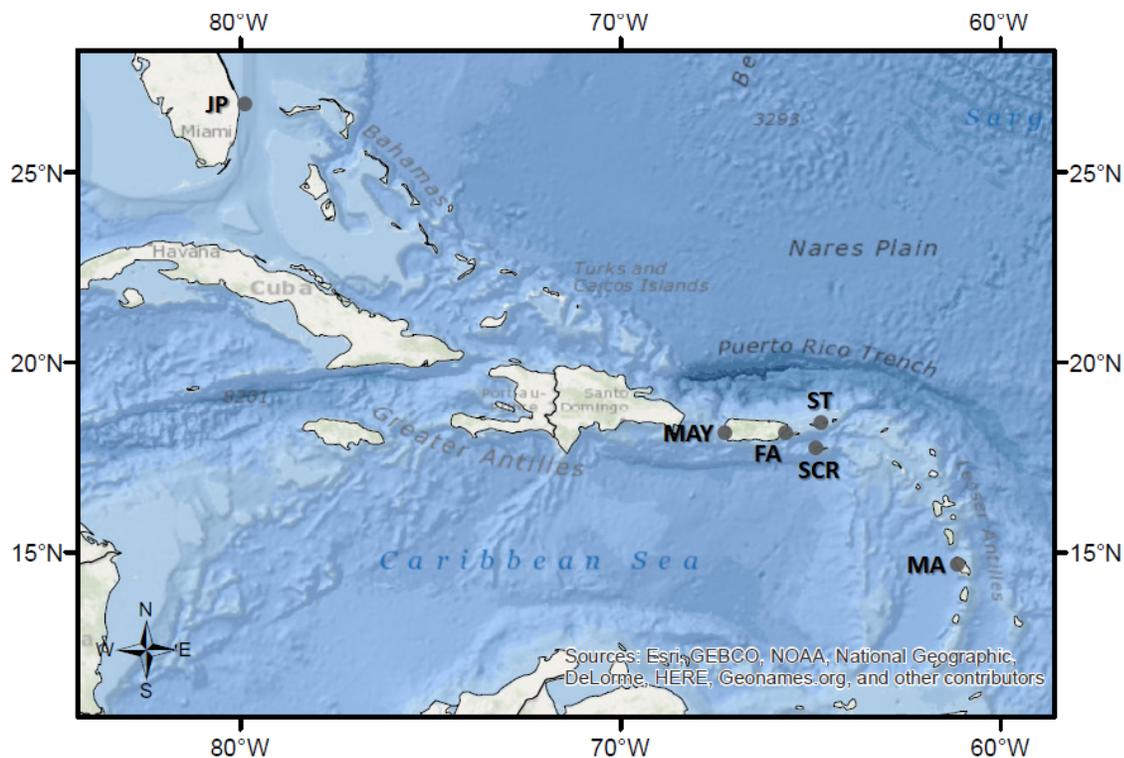


Figure 4. Queen triggerfish sampling localities

Sampling localities for the study of population structure in queen triggerfish. JP: Jupiter; MAY: Mayaguez, PR; FA: Fajardo, PR; SCR: Saint Croix; ST: Saint Thomas; MA: La Martinique.

2.2.2 Library preparation and sequencing

Genomic DNA was extracted using the phenol-chloroform method (Sambrook et al. 1989). DNA concentration for each individual sample was determined using a NanoDrop 2000

spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and the quality of the DNA was also evaluated during agarose electrophoresis. DNA solutions were concentrated or diluted in order to achieve a final concentration of 40-100 ng/ μ L. dd-RAD sequencing libraries were prepared from each individual DNA using protocols described in section 2.1.3.

2.2.3 Bioinformatics analysis of sequencing data and SNP Discovery

Sequencing reads were filtered as described in section 2.1.4 and processed in dDocent for mapping on the draft reference genome and SNP discovery. Only the genome contigs featuring SNPs genotyped in the linkage map experiment were used as a reference in order to increase the proportion of loci that could be ordered in the genome. SNPs were retained if their quality score was greater than 30, coverage was less than 67 (2.5 times the average depth of the dataset) and coverage was greater than 10 reads per sample. Two datasets were generated. One dataset was filtered to maximize the number of SNPs and the second aimed to maintain 50 or more individuals per locality sample. Filtration steps were performed in VCFTOOLS (Danacek et al. 2010) and PLINK v. 1.9 (Purcell et al. 2007).

Filtration for the High_SNP dataset allowed to 35% missing data per SNP, 25% missing data per individual with a total genotyping rate of 87%. Filtration for the High_sp_size allowed 25% missing data per individual and 15% missing data per SNP. The total genotyping rate in remaining individuals was 81.7%.

SNPs for which genotype frequencies departed significantly from Hardy-Weinberg equilibrium expectations during exact tests were discarded. A variant call file (VCF) including all the retained SNP genotypes was generated.

2.2.4 Data analysis

Data analysis began with outlier analyses to detect candidate loci undergoing divergent selection in different populations. Outlier detection employed two methods. First, OUTFLANK (Whitlock and Lotterhos 2015) was used with a q value threshold of 0.05 and retaining loci with a minimum heterozygosity of 0.01. The distribution of F_{ST} values was trimmed of the top 5% percentile before inference on outliers as recommended by Whitlock and Lotterhos (2015).

Outlier analyses was also conducted in PCAdapt (Luu et al. 2017). The parameter k representing the optimal number of population clusters for inference was determined based on the observation of the slope of the proportion of explained variance which supported selection of $k = 2$.

Based on the outcome of outlier analyses, all loci from the two datasets were processed for further analyses.

The software PGDSPIDER v.2.0.9.0 (Lischer and Excoffier 2012) was used to convert vcf files to plink format. Heterozygosity in each locality sample were computed in PLINK.

Weir and Cockerham's (1984) θ estimate of F_{ST} were obtained in the R package StAMPP (Pembleton et al. 2013). Pairwise F_{ST} values were plotted in a PCoA computed in ape 5.0 (Paradis and Schliep 2018).

A discriminant analysis of Principal Components DAPC (Jombart et al. 2010) was computed in the R package ADEGENET (Jombart and Ahmed 2011) where 120 principal components, comprising 55% of the observed variance in the data were retained as predictors for linear discrimination analysis.

A Mantel test of the correlation between genetic and geographic distance was computed in GENEPOP v.4.7 (Raymond and Rousset 1995, Rousset 2008) and the significance of the correlation was tested using 10,000 permutations. The genetic distance between samples (aggregated per locality for this analysis) was computed as $F_{ST}/(1-F_{ST})$ as recommended by Rousset (1997) where F_{ST} was the Weir and Cockerham (1984) θ . In this analysis, dispersal of queen triggerfish was assumed to follow a stepping stone model along the Antilles archipelagos where gene flow results from the transport of juveniles and larvae along the Antilles current in a one dimensional dispersal process. The geographic distance was calculated as a linear distance among sampling localities in kilometers along a transect that followed the Antilles archipelagos (Figure 5).

The slope of the regression between genetic and geographic distances (b) was used to calculate the parameter σ (the standard deviation of parental position relative to offspring position) using

the formula $\sigma = \sqrt{\frac{1}{4D_e b}}$ (equation 1) where D_e is the effective population density.

The effective population density used to estimate σ was derived as $D_e = \frac{N_{eff}}{L_T}$, where L_T is the length of the transect, expressed in kilometers, across the entire sampling range following the Antilles archipelago axis until Puerto Rico and then continuing to Florida along the Antilles current (Figure 5), and N_{eH} is the harmonic mean of the effective population size (N_e) estimates in the different localities.

Effective population size was estimated using both the linkage disequilibrium (LD) method (Hill 1981; Waples 2006) and the heterozygote excess (HE) method (Pudovkin et al. 1996) for the 6 sampling locations. Computations were performed using the software NEESTIMATOR v.2.0.1 (Do et al. 2014). Markers with minor allele frequencies below 0.05 were discarded in the calculation as recommended by Waples and Do (2010) considering the available sample sizes.

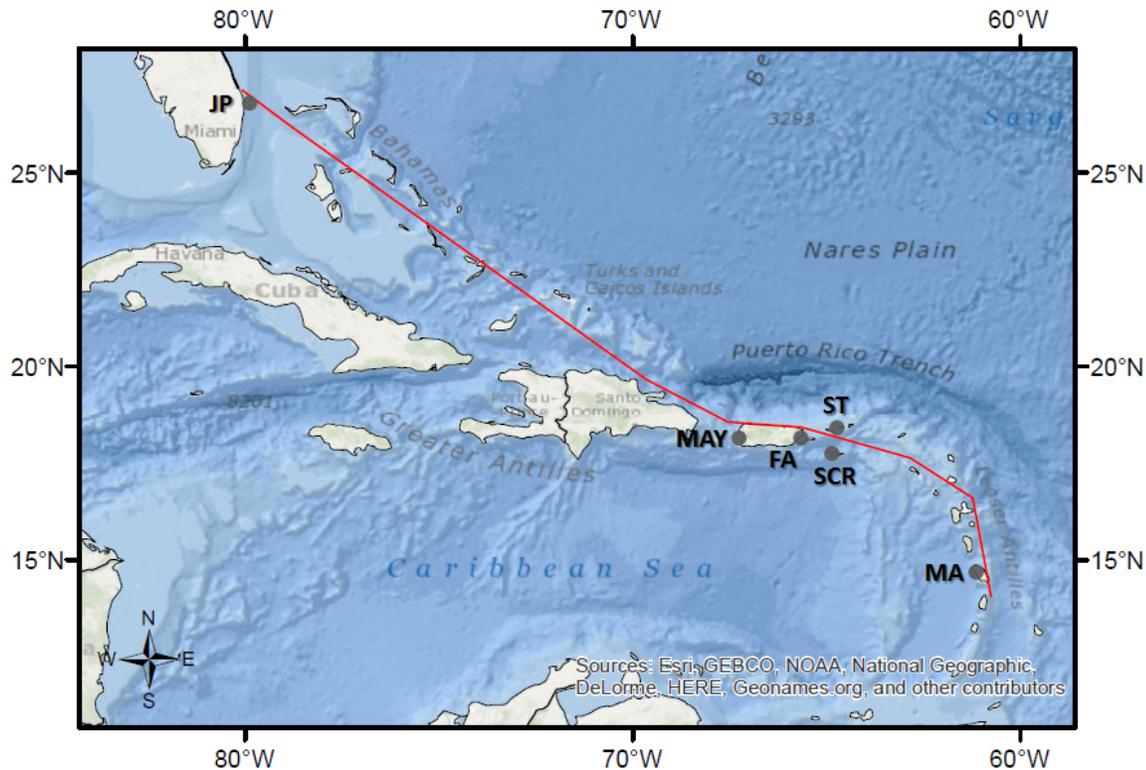


Figure 5. Transect line used to approximate the shortest distance between queen triggerfish geographic populations

Sampling locations for the study of population structure of queen triggerfish. The red line represents the transect used to describe the isolation-by-distance model.

LinkNe (Hollenbeck et al. 2016) was used to infer recent changes in effective population size in each locality sampled. N trajectory was estimated using a 5 cM bin size and the sliding window for calculating moving averages was set at 1 cM. Alleles with frequency less than 5% were omitted for computations.

Relatedness analysis was conducted in the R package ‘related’ (Pew et al. 2015). Initially, 100 related and unrelated dyads were simulated using allele frequencies from the dataset and 500 loci were employed to test the effectiveness of 4 different moment relatedness estimators (Queller and goodnight 1989, Li et al. 1993, Lynch and Ritland 1999, Wang 2002). All estimators returned correlations between simulated and estimated relatedness over 90%. The highest accuracy was obtained using the Lynch and Ritland (1989) estimate (correlation over 94%) which was selected for further analysis. Pairwise relatedness analysis was computed between all samples and 1,000 bootstrap samples were used to estimate 95% Confidence Intervals for the obtained coefficient.

Structure was also examined based on individual multilocus genotype data by constructing a minimum spanning tree of genotypic differences, in the R package poppr. The first two principal components from PCA analysis were used to compute Euclidean distances between the genotypes.

3. Findings

3.1 Draft reference genome

3.1.1 Genome sequencing and assembly

Sequencing yielded 265,984,922 paired end reads (151 bp read length) for a total of 80,327,446,444 bp total sequencing. The total length of the assembly was over 778 Mb but there were over 4.1M contigs and a large fraction of the assembly was in small contigs with only 344,125,808 bp in contigs over 1,000 bp. The N50 and L50 were 1,515 bp and 100,353 bp respectively.

The k-mer counting method using a k-mer size of 18 yielded an estimate of genome size of 574,021,784 bp.

3.1.2 Linkage mapping

The mean number of reads per larval sample before filtering averaged 7,450,187 and ranged between 1,132 and 21,232,757. The distribution of pairwise relatedness values among larvae samples was uni-modal and consistent with all larvae belonging to a single full sib. Linkage mapping proceeded with the 184 individuals for which sequencing depth was greater than 1.5M reads. The mean number of reads per sample after filtration and removal of PCR duplicates was 3,206,425.

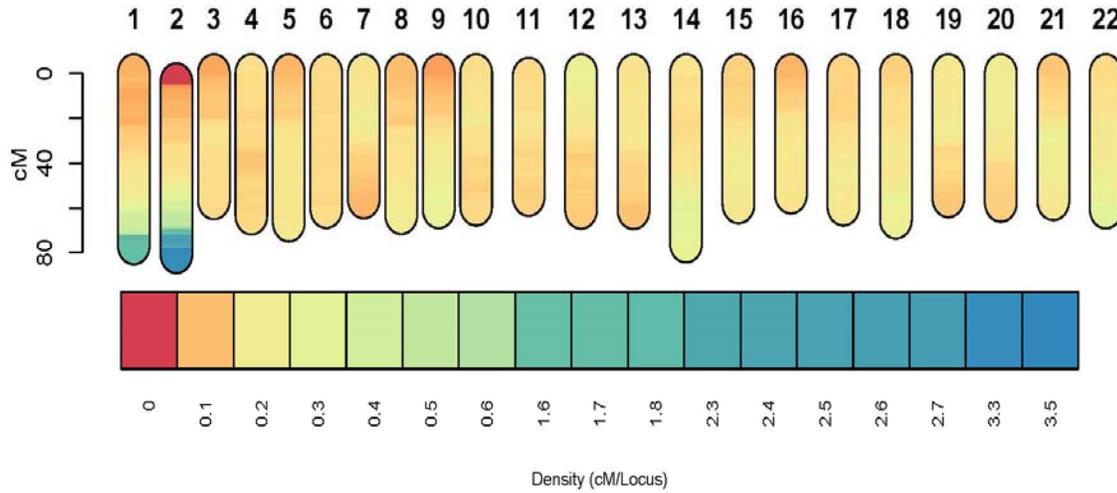
dDocent yielded 7,550,656 loci of which 15,027 were assigned to one of 22 linkage groups in Lep-MAP. Loci were ordered in a single average map reflecting recombination events in the two parents (the male and female parent were not sampled which prevented description of sex specific maps). The map spans 1,365.37 cM and anchors 10,367 genome contigs corresponding to a total anchored length of 19,95 Mb. The characteristics of each linkage group (numbers of markers and length in cM calculated using the Kosambi function) are presented in Table 1.

Table 1. Map length in cM and number of markers assigned to each of the 22 linkage groups of the *Balistes vetula* linkage map

<i>linkage group</i>	<i>cM</i>	<i># loci</i>	<i>linkage group</i>	<i>cM</i>	<i># loci</i>
1	76.87	934	12	60.61	663
2	80.81	866	13	60.83	680
3	56.51	845	14	76.07	624
4	66.66	833	15	58.16	603
5	66.32	774	16	53.76	639
6	60.26	756	17	59.16	620
7	56.54	669	18	64.99	575
8	63.04	727	19	55.45	580
9	60.40	699	20	57.52	575
10	59.21	676	21	56.76	527
11	54.92	688	22	60.52	474
			Total	1365.37	15,027

The marker density was 0.2 cM/locus or less for most of the map (average 0.09 cM/locus) except for the ends of LGs1 and 2 that had both regions of low density 1.6-3.5 cM/locus (Figure 6).

Population genetic analyses focused on contigs anchored by the linkage map to enable accounting for physical linkage during inference of genomic regions under selection and estimation of effective population size trajectories.



Rendered by LinkageMapView

Figure 6. Chart representing the marker density on each of the 22 linkage groups inferred for queen triggerfish.

3.2 Population genomic survey

Sequencing of the libraries on the Illumina platform yielded an average of 832,329 to 1,174,495 filtered reads per sample. Within these reads a total of 1,251,788 loci were identified in FREEBAYES.

Table 2. Numbers filtered reads per locality sample

# <i>reads/spl</i>	<i>JP1</i>	<i>JP2</i>	<i>MAY</i>	<i>FA</i>	<i>ST</i>	<i>SCR</i>	<i>MA</i>
Mean	948,925	993,925	1,168,042	968,474	1,174,495	832,329	964,700
Max	2,485,332	1,554,142	2,279,799	1,761,966	1,862,520	1,743,911	2,279,799
Min	209,086	168,316	210,799	275,627	212,745	212,634	168,406

The High_SNP filtration strategy yielded 4,512 loci (average 205.1 per linkage group) with a genotyping rate of 86.9%. The numbers of individuals retained per population averaged 45.4 but samples sizes were low for JP2, MA and SCR (24, 34 and 35 respectively).

The High_sp_size filtration yielded 1,698 loci (average 77.2 per linkage group) and on average 74.6 individuals per sample (range 54-86, Table 3).

Table 3. Sample sizes retained for the High_SNP and High_sp_size datasets

	<i>JP1</i>	<i>JP2</i>	<i>MAY</i>	<i>FA</i>	<i>ST</i>	<i>SCR</i>	<i>MA</i>
High_SNP	24	78	45	46	56	35	34
High_sp_size	72	79	73	77	86	72	54

Hierarchical clustering of all genotypes revealed occurrence of 22 individuals highly divergent from the rest of the samples in the High_SNP dataset (Appendix 1). Because such individuals may represent artifacts of genotyping or sampling such as tissue contamination, these samples were removed from analysis.

OutFLANK detected 12 outliers at a q-value threshold of 0.05. The F_{ST} values of outliers ranged between 0.053 and 0.151. Only two loci had F_{ST} values above 0.1. PCAdapt detected 16 outliers, also at a q-value threshold of 0.05, but none of the outliers in this panel overlapped with the OutFLANK outliers (Figure 7 a and b) overall giving weak support for outlier regions in the dataset. Similar results were obtained with the High_sp_size dataset (not shown).

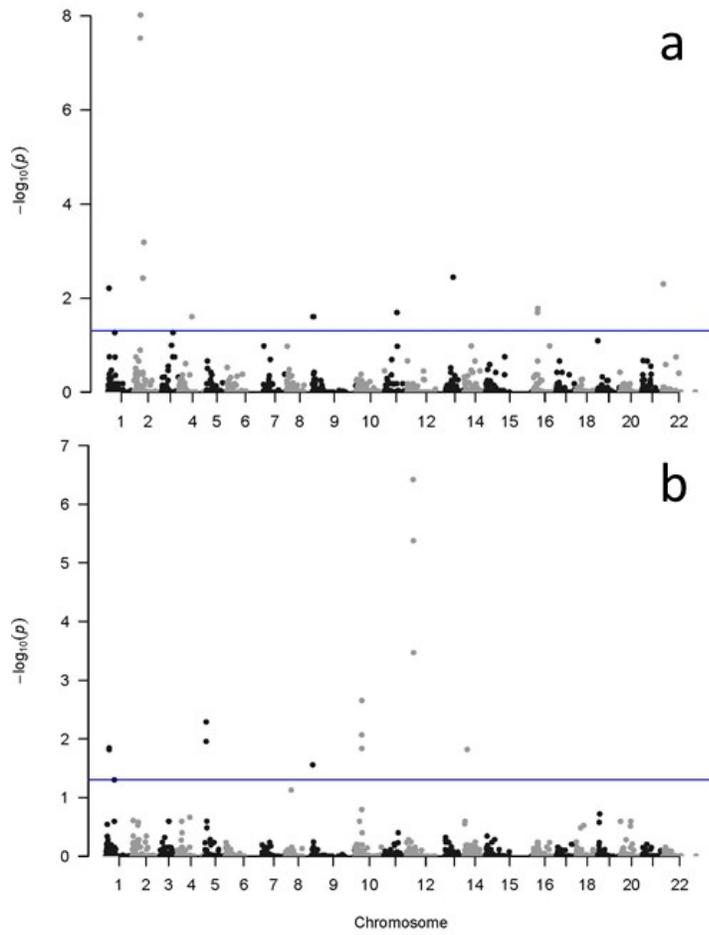


Figure 7. Manhattan plot of outlier loci detected in OutFLANK (a) and PCadapt (b)

Heterozygosity distributions were similar among samples (Figure 8).

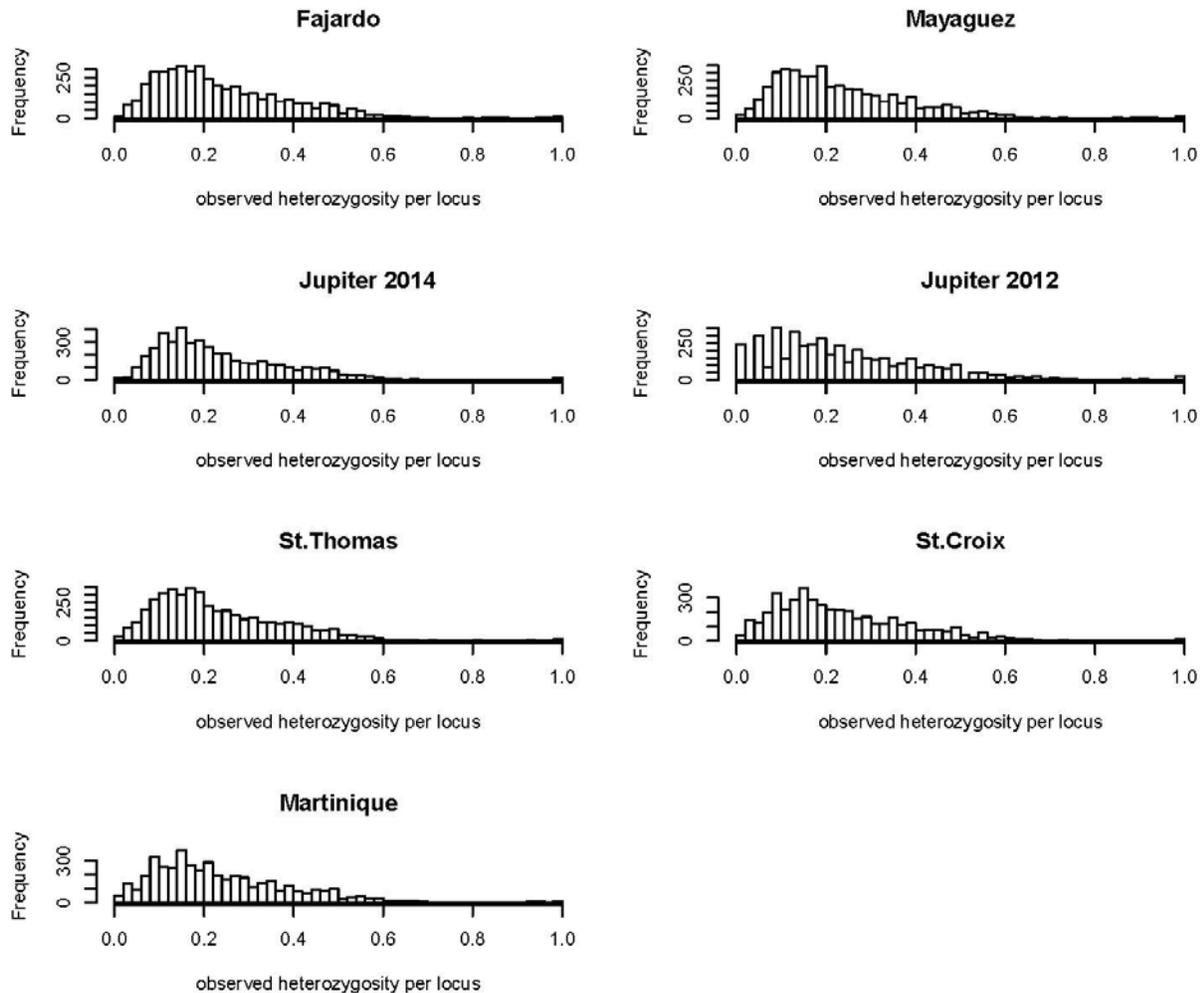


Figure 8. Distribution of per locus heterozygosity in 7 samples of queen triggerfish from 6 localities

Pairwise F_{ST} estimates obtained from the two datasets were very low (range -0.00014 - 0.0083 for the High_sp_size and 0 - 0.0009 for the High_SNP datasets respectively). PCoA of the High_sp_size dataset showed that the two Jupiter samples diverged from the others (Figure 9a) and from each other as well, a trend that was supported statistically during exact tests of population differentiation (Table 4a). The sample from la Martinique also appeared to diverge from other samples although F_{ST} estimates were lower and uncorrected probability values were

close to 5%. The pattern of structure was less apparent in the High_SNP dataset where most pairwise comparisons gave non- significant outcomes (Figures 9b, Table 3b).

Table 4a. Pairwise F_{ST} estimates and P-values during homogeneity tests comparing samples of queen triggerfish geographic populations (High_sp_size dataset)

JP1	JP2	FA	MAY	ST	SCR	MA	
	<0.001	<0.001	<0.001	0.003	<0.001	<0.001	JP1
0.0070		<0.001	<0.001	<0.001	<0.001	<0.001	JP2
0.0048	0.0019		0.337	0.071	0.659	0.006	FA
0.0045	0.0015	0.0001		0.105	0.482	0.112	MAY
0.0052	9.3E-05	0.0005	0.0004		0.092	0.053	ST
0.0047	0.0013	-0.00014	6.2E-05	0.0006		0.058	SCR
0.0053	0.0018	0.0011	0.0005	0.0083	0.0073		MA

Estimates of F_{ST} (lower diagonal) and exact probability (upper diagonal) obtained during pairwise homogeneity tests comparing queen triggerfish geographic samples.

Table 4b. Pairwise F_{ST} estimates and P-values during homogeneity tests comparing samples from queen triggerfish geographic populations (High_SNP dataset)

JP1	JP2	FA	MAY	ST	SCR	MA	
	0.441	0.487	0.24	0.122	0.349	0.326	JP1
8.20E-05		0.016	0.478	0.012	0.230	0.215	JP2
8.90E-06	0.0006		0.73	0.217	0.315	0.114	FA
0.0005	2.20E-05	0		0.204	0.884	0.123	MAY
0.0009	0.0006	0.0002	0.0006		0.232	0.049	ST
0.0003	0.0003	0.0002	0	0.0003		0.769	SCR
0.0003	0.0003	0.0005	0.0004	0.0007	0		MA

Estimates of F_{ST} (lower diagonal) and exact probability (upper diagonal) obtained during pairwise homogeneity tests comparing queen triggerfish geographic samples.

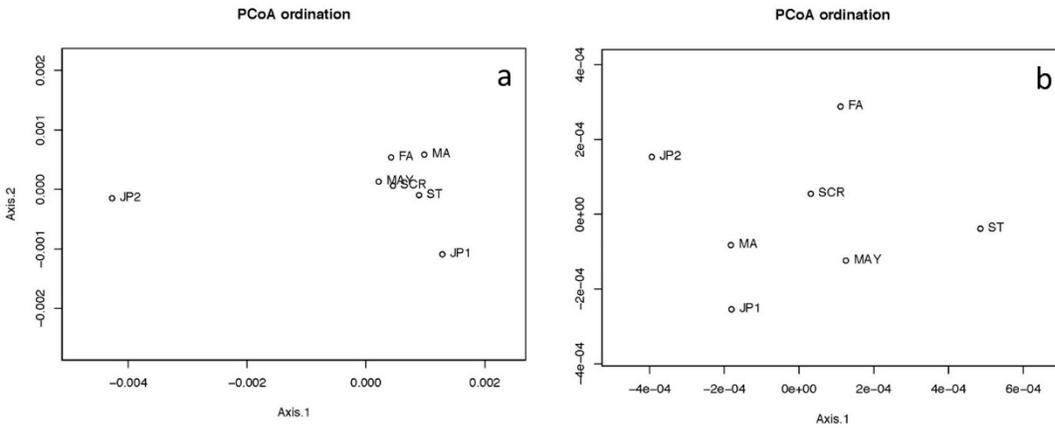


Figure 9. PCoA ordination of pairwise F_{ST} obtained during comparison of 7 queen triggerfish samples from 6 geographic locations. a High_sp_size dataset, b High_SNP dataset

The discriminant analysis of principal components also indicated weak patterns of structure with a trend for the two samples from South Florida to diverge from each other and from the Caribbean samples, consistent with the PCoA ordination of F_{ST} estimates (Figure 10).

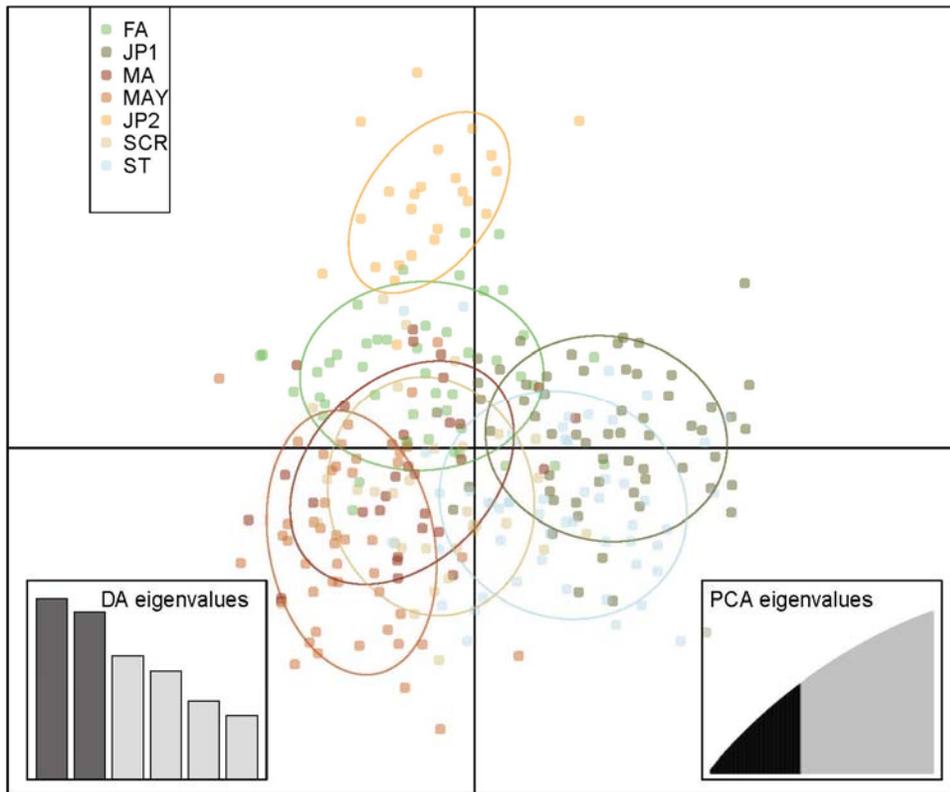


Figure 10. Discriminant analysis of principal components of the High_SNP dataset (similar results were obtained for the High_sp_size dataset not shown).

Pairwise relatedness estimates within locality showed a distribution centered around zero and suggested occurrence of a few closely related individuals in most samples (Appendix 2). The minimum spanning network of genotypes did not reveal clustering of genotypes in association with geographic origin (Appendix 3). The aggregation of relatedness coefficient across localities gave the same outcome.

Because the samples from South Florida obtained in 2012 and 2014 differed significantly from each other, estimates of the slope of isolation by distance were generated separately accounting for each of the two temporal samples. The probability of significance of the correlation between genetic and geographic distance during the Mantel test were 0.014 and 0.224 when JP1 and JP2 were used, respectively. Point estimates and confidence interval for the slopes were 2.6×10^{-6} (95% CI $1.5 \times 10^{-6} - 4.4 \times 10^{-6}$) using the JP1 sample and 2.8×10^{-7} ($2.2 \times 10^{-8} - 5.5 \times 10^{-7}$) using JP2

(Figure 11). The corresponding estimates of σ were 728 (558 - 968) using JP1 and 2,223 (1,578 - 7,874) using JP2.

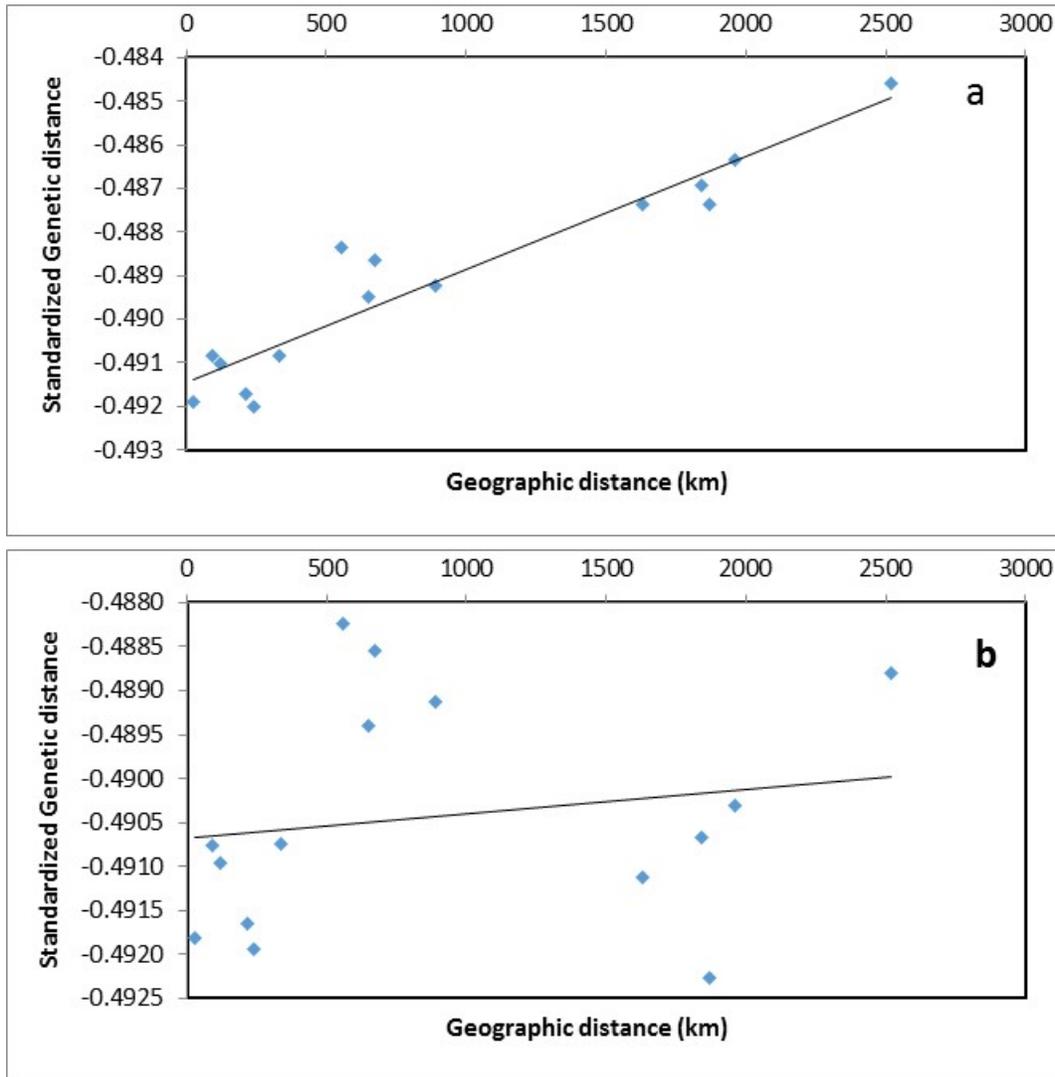


Figure 11. Relationship between genetic and geographic distance in 6 geographic populations of queen triggerfish. a) using JP1 and b) using JP2 to compute genetic distances between localities.

Estimates of effective population size by the heterozygote excess method were all infinite.

Estimates by the linkage disequilibrium method are presented in Table 5. The lowest values were

obtained for the JP2 and MA samples (point estimates were 98 and 230 respectively) and intermediate values were also obtained for JP1 and SCR (point estimates were 924 and 476 respectively). Most Confidence Intervals had infinite upper bound likely due to the relatively small sample sizes available. However, estimates of N_e in the recent past suggested a clear decline within the past 20 generations for all localities with a trend of moderate increase for the US Caribbean localities during the last generation prior to sampling (Figure 12).

Because isolation by distance integrates the effects of genetic drift over multiple recent generations, the harmonic mean of past N_e obtained from LinkNe for all localities was used to compute a harmonic mean of N_e across the region (across localities) which was divided by the length of the transect to estimate population density. The harmonic mean of N_e was 456 and corresponded to an effective density D_e of 0.181 which was used to generate the estimates of sigma reported above based on equation 1.

Table 5. Estimates of effective population size by the linkage disequilibrium method for 6 queen triggerfish geographic populations (two temporal estimates for the JP locality, JP1 and JP2)

<i>Locality</i>	<i>Harmonic mean sample sizes</i>	<i>N_e (95% CI)</i>
JP1	48.3	924 (177 – infinite)
JP2	71.6	98 (55 –270)
FA	52	40,382 (192 - infinite)
MAY	55.5	Infinite (278 - infinite)
ST	62.4	Infinite (356 - infinite)
SCR	41.2	476 (74 - infinite)
MA	36.2	230 (45 – infinite)

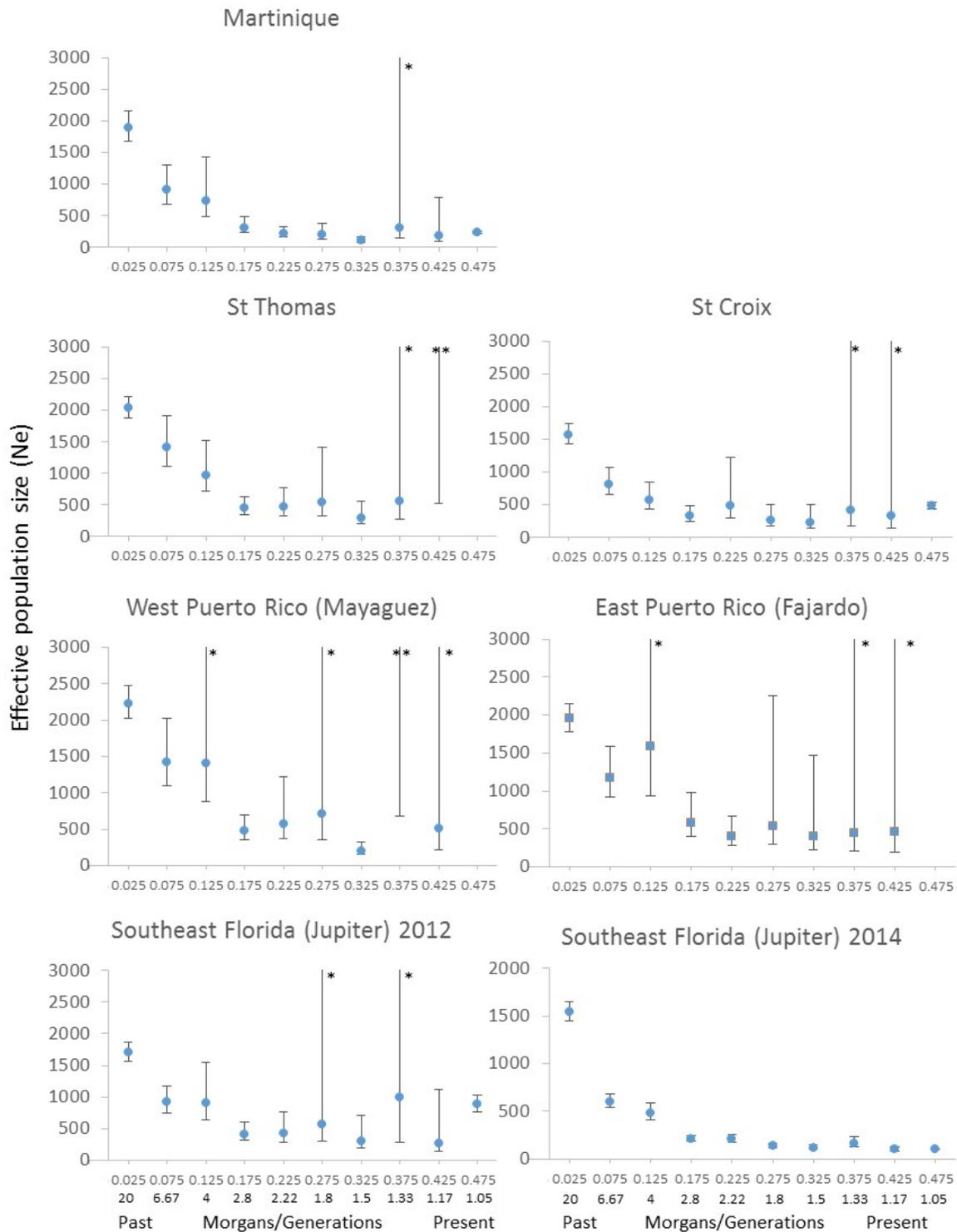


Figure 12. Effective population size trajectories in recent generations as estimated from LinkNe
 * Upper bound of confidence Interval above scale, ** Point estimate and upper bound of Confidence Interval above scale

4. Interpretations

The project proposal initially involved characterizing queen triggerfish populations using microsatellite markers but the recent improvement of the cost effectiveness of genotyping by sequencing methods made it possible to incorporate high-density genome scans to study non-model species such as the queen triggerfish. Therefore the project team decided to revise the workplan and implement this method as indicated in progress reports. The dd-RAD sequencing method employed in this work yielded much more information on genomic variation in queen triggerfish than would have been achieved with the microsatellites (up to 4,512 genetic markers were surveyed instead of at most 20 microsatellites intended in the original project). Because efforts to obtain queen triggerfish samples from some of the Caribbean locations targeted for the study were unsuccessful and the opportunity to obtain samples needed to develop a linkage map of the queen triggerfish genome did present, the project team decided to focus the end of the project on developing genomic resources (a draft genome sequence assembly and a high density linkage map) for the species. These resources were then applied to improve the population genetic inferences that could be made by allowing in particular examining recent trajectories in effective population size.

4.1 Genomic resources

The first objective of this project was to develop genomic resources to improve interpretations of genome scans during population genomic studies of the queen triggerfish in this project and future genetic studies. The project funding was insufficient to produce a full draft reference genome but a first assembly was generated using Illumina short sequencing reads which could be obtained in sufficient numbers to cover reliably the queen triggerfish genome for moderate costs. The obtained assembly is fragmented (over 4.1 M contigs total) but the total length in contigs is close to the estimate of the size of the queen triggerfish genome obtained in this study (575 Mb) and consistent with c-values reported for triggerfish (0.72 pg, Hinegardner and Rosen 1972). The utility of the reference genome was improved by anchoring genome contigs to a high density linkage map. The New England aquarium provided samples of queen triggerfish larvae from a spontaneous spawning event that occurred in one of the aquarium tanks. The larvae that were

sampled were confirmed to belong to a single full sib and a sufficient number of offspring could be amplified and processed for sequencing to develop a high-density linkage map.

Linkage analysis led to assign loci to 22 linkage groups, a number consistent with the number of chromosomes reported in the karyotype of the queen triggerfish (Sá-Gabriel and Molina 2005). The map includes 15,027 loci and spans over 1,367 cM yielding an average inter marker interval of only 0.09 cM. Considering the genome size estimate obtained in this study (575Mb), this average inter marker interval corresponds to a physical interval of slightly over 38kb per locus, a density that is expected to allow capturing a significant fraction of genomic regions subjected to divergent selection. A large fraction of the contigs obtained during assembly are not anchored by the linkage map: Only 10,367 contigs of the 4.1M present in the assembly spanning a total length of 19,95 Mb (i.e. approximately 3.4% of the queen triggerfish genome) are anchored by the map. However, the use of a reduced reference sequence consisting of contigs containing SNPs discovered in the mapping family yielded 4,512 and 1,698 in the High_SNP dataset and the High_sp_size dataset respectively. These datasets enabled a first investigation of outlier genomic regions and a demographic analysis of recent effective population size trajectories and robust estimates of genomic variation. Future efforts to develop genomic resources in queen triggerfish should focus on obtaining long sequencing reads (e.g, using the PACBIO Sequel platform) which will allow reducing the reference assembly to a few thousand contigs or less and, with the help of the linkage map developed in this study, tracking comprehensively genome wide variation and the genetic bases of local adaptations.

4.2 Stock structure and demography

Samples could be obtained from four localities in the US Caribbean, Southeast Florida in the continental US and from La Martinique Island in the lesser Antilles. Two strategies were used to filter sequencing reads obtained during sequencing. A first dataset (High_SNP strategy) maximized the number of loci recovered (4,512 instead of 1,698) but at the expense of sample sizes. Increasing the number of loci was meant in particular to increase the density of the genome scan during outlier analyses to detect regions of the genome undergoing divergent selection. The second dataset aimed to recover a high fraction of the individuals genotyped in each of the 7

locality-samples and maximize power during analyses of demographic parameters and neutral structure which are less demanding in scan density. The first step of the analysis tested whether some of the genetic loci and genomic regions were impacted by divergent selection and local adaptation through an outlier analysis. Analyses using OutFLANK and PCAdapt yielded very low numbers of candidate outliers (12 and 16, respectively) and none of the significant loci identified overlapped between the two methods giving no reliable support for loci or genomic regions undergoing divergent selection.

Considering the large number of loci surveyed in this genome scan (4,512 in the High_SNP dataset), the estimates of genome size (575 Mb) and the length of the map (1,367 cM), loci under selection would be expected to be located on average within 0.3 cM or 127kb of one of the genetic loci surveyed in this study. In this situation, the signature of local adaptation would be expected to yield a F_{ST} signal at some of the SNPs (framing selected loci) in a broad range of demographic scenarios (Charlesworth et al. 1997, Storz 2005). Thus, the lack of significant outlier in the present study suggests that no genomic region is undergoing strong divergent selection although occurrence of genomic regions affected by moderate selection cannot be rejected. Previous studies of marine species with large dispersal capabilities using high-density genomic scans revealed the occurrence of outliers in association with significant neutral population structure (Nielsen et al. 2009, Bradbury et al. 2010, Limborg et al. 2010, Laconcha et al. 2015), but also in cases where there was no significant spatial structure at neutral loci (Lamichhaney et al. 2012, Grewe et al. 2015). The present study on queen triggerfish thus contrasts with these studies in that no significant outliers were detected. However, as pointed out by Lotterhos and Whitlock (2014), several of the methods used to detect outliers have a high risk of false positives, especially when no structure is present at neutral loci. In consequence, some of the studies reporting outliers, in particular those where outliers were not found associated with detectable neutral structure, may in fact have been impacted by high rates of false positives and thus situations similar to the present study on queen triggerfish with no significant support for outlier loci may be more common than currently apparent in the literature. Considering that queen triggerfish have a high potential for gene flow, it is possible that the lack of differentiation is the result of a selection-migration balance where the differentiation caused by divergent selection and local adaptation is counterbalanced by gene flow (gene swamping, Lenormand 2002, Conover et al. 2005, Cheviron and Brumfield 2009). A first hypothesis would be that

queen triggerfish are only found in habitats with similar characteristics leading to little or no local selection. A second hypothesis is that local selective pressure is present but migrations are sufficient to prevent divergence at impacted loci. As pointed out by Lenormand (2002), even though the potential for adaptation is greater in sparsely populated environments such as those occupied by queen triggerfish outside their center of abundance, the homogenizing effects of gene flow is also stronger in those populations with a higher rate of effective immigration from the larger stocks in the center of abundance of the species. Such a scenario is plausible in the range surveyed in this study considering the high dispersal distance estimates compatible with direct transport of migrants across the sampling surface. Therefore, if local selective pressures are important determinants of the fitness of queen triggerfish populations, it would be important to rebuild healthy local stocks with large effective size so that the impact of migrants would be reduced and the selection of genotypes with higher fitness becomes more efficient.

Based on the negative outcome of outlier tests, the analysis of spatial genetic variation and demographic parameters continued assuming all the loci were evolving neutrally. Divergence among geographic samples was very low ($F_{ST} < 0.0083$ for the High_sp_size dataset and < 0.0009 for the High_SNP dataset) and isolated primarily the two samples from South Florida which were also different from each other. To a lesser degree, the sample from La Martinique also diverged from the other geographic samples. A weak isolation-by-distance pattern was suggested by a positive correlation between genetic and geographic distance although the slope of the relationship was only significant in one of the tests performed. Studies of population structure in the Southeast Florida and US Caribbean were recently conducted in three Lutjanids, sharing the adult sedentary behavior and pelagic larval lifestyle of the queen triggerfish. Although these species are characterized by a shorter pelagic phase (30-40 days, Lindenman et al. 2000) than that of the queen triggerfish, divergence among geographic samples was also very weak for all three snappers ($-0.009 - 0.0095$, Carson et al. 2011; Gold et al. 2011; Saillant et al. 2012).

The observation of very weak levels of divergence and lack of statistical significance of homogeneity tests in this study are thus consistent with the low levels of divergence reported in all three snapper species and the prediction of dispersal across longer distances based on the longer larval pelagic phase (almost twice the duration of that of snappers). Opportunities to

overcome barriers such as the separation between the two Puerto Rican platforms or the longer distance between the US Caribbean and South Florida are much greater in queen triggerfish. Homogeneity in the frequency of alleles was indeed observed even when the population from La Martinique was accounted for, bringing a total geographic distance covered by the sampling design to 2,518 km.

Divergence between the northern and Southern Puerto Rican platforms and a possible isolation of populations located West of Puerto Rico were detected in yellowtail snapper (Saillant et al. 2012), and, for both yellowtail snapper and lane snapper, divergence between the US Caribbean populations and South Florida was inferred (Gold et al. 2011, Saillant et al. 2012). These results possibly reflected an isolation-by-distance pattern as hypothesized in this study for queen triggerfish. Also, in the mutton snapper, differences in effective population size estimates among regions of the Caribbean suggested demographic independence of nearby populations, despite of the lack of significant divergence in allele frequencies in this species (Carson et al. 2012). These results suggest that structure and demographic independence may occur on a small geographic scale, possibly due to physical factors such as the separation of the two Puerto Rican platforms, differing levels of local retention, or differing exploitation rates among regions, even when divergence in allele frequencies is very weak. The study of effective population size did provide further insights into the degree of demographic independence of local queen triggerfish populations as in the case of mutton snapper (Carson et al. 2012).

In the queen triggerfish, even though gene diversity estimates did not differ among sampling locations, estimates of effective population size indeed suggested differences among regions with low estimates for South Florida (harmonic mean of the two samples 177), St. Croix (476), and La Martinique (230) while the remaining samples (East and West of Puerto Rico and St Thomas), had much larger point-estimates (40,382 or infinite). Confidence intervals were large and mostly overlapping but estimates of contemporaneous effective population size were highest in the northern Puerto Rican platform where the species is currently exploited most intensively and thought to be most abundant. Analysis of recent effective population size trajectories however indicated a reduction of N_e during the past 20 generations in all localities. Considering the high connectivity that was inferred across the surveyed range, these results may signal a recent decline of queen triggerfish across the region. However, inferences on the past trajectories in effective population size need to be taken with caution considering the life history of queen

triggerfish and the potential for admixture from different spawning assemblages each generation. If triggerfish colonizing habitats originate from different mixtures of spawning stocks each generation due to specific larval transport conditions, the resulting complex scenarios of admixture could alter the linkage disequilibrium signal and the estimated values of past N_e . The impacts of such migration and admixture events on estimates produced by LinkNe still need to be formally evaluated (Hollenbeck et al. 2016). Assuming the demographic signal inferred during this analysis is robust to this departure from assumptions made in the analysis, all geographic stocks would have declined substantially during recent generations with one or more point-estimates of N_e below 500 during the past 3 generations. Estimates were in some cases substantially lower than 500 (see e.g. estimates for the Jupiter 2014 sample).

According to the “50/500 rule” defined by Rieman and Allendorf (2001), an effective population size of at least 50 is sufficient to minimize inbreeding effects, while N_e greater than 500 would be necessary to maintain the equilibrium between the loss of adaptive genetic variation from genetic drift and its replacement by mutation. Accordingly, an effective size greater than 500 is often considered a minimum target for management of populations in order to achieve long-term sustainability. Based on the results from this study, localities on the northern Puerto Rican platform had sufficient effective population sizes to prevent immediate risk of extinction due to genetic factors in the breeding pools that produced the 2014 samples. However, the analysis of recent trajectories for N_e leads to a different interpretation on their status because effective population size would have been once or several times close to or below the 500 threshold in recent years in all localities. The recent bottlenecks inferred in all localities suggests that N_e would be close to 500 or below this value across the area because genetic diversity maintained in a population approaches the harmonic mean of N_e over time. Thus, despite apparent healthy demographic stocks in most of the US Caribbean, queen triggerfish may be at risk other the long term.

The southern Puerto Rican platform (St. Croix sampling location) had a lower effective size (and effective population size trajectory) than localities on the northern platform suggesting demographic independence of the two stocks despite geographic proximity (less than 200 miles for West Puerto Rico, less than 70 miles for St Thomas). Swearer et al. (1999) found that bluehead wrasse (*Thalassoma bifasciatum*), a reef fish with a long planktonic larval duration, exhibit recruitment around St. Croix dominated by local retention. This is particularly true from

June to August (Swearer et al. 1999). One of the queen triggerfish peak spawning periods occurs between August and October (Aiken 1983). Even though this only partially overlaps with the period during which local recruitment seems to be favored, increased local retention, at least at the beginning of the spawning season, could contribute to the lower effective population size estimate around St. Croix. However, considering the geographic proximity with the other upper Caribbean sampling locations where estimates of N_e were much larger, the low effective size in St. Croix may also reflect a temporal artifact and this result needs to be confirmed using additional samples.

The estimate of effective population size for the South Florida region based on samples obtained in 2014 was very small (98) which led to the genotyping of another sample (collected in 2012) to determine if the very low effective size was a temporal artifact. The estimate obtained for the 2012 sample was larger (point estimate 924) and the sample from 2012 diverged significantly from the one obtained in 2014 both in allele frequencies and in past N_e trajectory. Considering that genetic diversity and past (20 generations ago) effective population size were similar and relatively high in both samples, a possible explanation to these results could be that a bottleneck occurred in the 2014 generation or the 2014 generation was affected by a sweepstake pattern of recruitment where only a limited fraction of the spawning stock actually contributed to offspring (Hedgecock and Pudovkin 2011). Another possible explanation is that queen triggerfish settling in South Florida habitat originate largely from Caribbean spawning aggregations rather than local spawning biomass. Under this hypothesis, South Florida habitats would be colonized by occasional larval pulses from Caribbean locations when pelagic transport conditions are favorable. This hypothesis would be consistent with the variable and sometimes small effective population size estimate in that recruits in different years could descend from different Caribbean spawning aggregations, whose offspring reached South Florida that particular year. Spawning aggregations of queen triggerfish appear to be small (Heyman et al. 2013; < 40 specimens, R. Nemeth, University of the Virgin Islands Personal communication) and would be expected to produce pools of migrants of limited effective size, a scenario consistent with the observation of a small N_e estimate in one of our sample but relative genetic homogeneity between the Florida samples and the Caribbean populations. Insights on the occurrence and status of a breeding population in South Florida could be gained by looking for spawning aggregations or assessing if adults are spawning capable or actively spawning during the spawning season.

Connectivity across the area was further examined by estimating the dispersal parameters using genetic data and the inferred isolation-by-distance relationship. Dispersal parameters are directly impacted by the population density and, in this study, this quantity was derived from the harmonic mean of the N_e estimates obtained for individual sampling localities. Because estimates of allele frequencies for the South Florida population and estimates of pairwise F_{ST} differed when the 2012 and 2014 samples were used, two estimates of the isolation by distance parameters were generated. The obtained slopes yielded point-estimates of the parameter σ of 728 and 2,223 using JP1 and JP2 respectively corresponding to a mean dispersal distance of 581 and 1,773 km as calculated assuming a normal dispersal distribution function in one dimension (Puebla et al. 2012). The lower bounds of the 95% CI corresponded to mean dispersal distances at 445 and 1,259 km respectively. While the two estimates generated in this study are quite different, they are both consistent with dispersal over very long distances. Considering our sampling design, the estimates reflect for a large part dispersal promoted by the Antilles current system. According to calculations based on the average velocity of this current, and considering a pelagic phase lasting up to 63-83 days, an average dispersal distance along the Antilles chain of 218-645 km was predicted, a prediction consistent with the genetic estimate obtained accounting for the 2012 sample from Jupiter. The sampling surface spanned slightly more than the average estimated dispersal distance (JP2 estimate) or 3 times this average for the JP1 estimate. In consequence, individuals dispersing at long distances were not expected to be captured by the sampling design, leading to a potential underestimation of sigma (Leblois et al. 2003). Sampling additional populations to expand the size of the lattice used to derive estimates would be beneficial to estimate more accurately dispersal parameters. Efforts to obtain samples from the Bahamas or in the lesser Antilles South of La Martinique as needed to expand the length of the lattice were however unsuccessful in this study. Another caveat of the sampling design is the possible occurrence of a barrier to migration between the Bahamas and South Florida due to the circulation patterns in that region and the possible transport of queen triggerfish larvae from the southern Caribbean to South Florida through the Caribbean-loop current system instead of the Antilles current. This would contribute to homogenize South Florida and Caribbean populations because the Caribbean current is fast and potentially alter the slope of the isolation by distance model which was otherwise influenced primarily by dispersal through the Antilles current. Sampling in the Southern Caribbean and other geographic regions in the path of this current (Belize) was unsuccessful but determining the

extent of populations along the path of this system and their potential contributions to South Florida would be useful to determine connectivity along this axis. Accordingly, samples from the Bahamas would be better suited to characterize the northern tip of the Antilles and estimate dispersal parameters along the Antilles current axis than South Florida. Samples from the Bahamas would also enable assessing whether gene flow between South Florida and the northern Antilles is restricted as hypothesized in this work.

4.3 Conclusions

The main objective of this work was to assess genetic stock structure of queen triggerfish in US waters and the connectivity of US stocks with other geographic stocks. While allele frequencies appear relatively homogeneous across the surveyed area, estimates of effective population size and of their recent trajectories vary among geographic regions. This could be due to some degree of demographic independence of local stocks but also reflect patchiness of recruitment of larvae dispersing from several spawning assemblages across the Caribbean in a highly connected metapopulation. The estimates of dispersal distance distributions indicate that recruitment in US populations could include contributions from spawning stocks located anywhere in the Caribbean region or the northern part of South America, highlighting the importance of maintaining healthy populations across the range in order to conserve this species.

The large volume of catches in the US Caribbean along with the rarity of queen triggerfish in other parts of the distribution range suggest that the northern Caribbean populations (including the US Caribbean) may represent the center of abundance of the species. Considering the range-wide perceived decline of queen triggerfish (except in the northern Caribbean region), determining the connectivity of the northern Caribbean stock with other remaining populations is essential in order to identify geographic areas that are unlikely to be replenished through migrations from this area and focus conservation efforts on their local spawning stocks. Based on circulation patterns in the region, potential sources of migrants to the Caribbean would be located in South and Central America. Considering the low abundance of queen triggerfish in those regions, the contribution of these stocks to recruitment and genetic diversity in the Caribbean is likely limited. In addition, the orientation of surface currents mostly South to North predicts the lack of gene flow from the northern Caribbean to southern populations in the lesser

Antilles or South America. These southern populations therefore require specific conservation efforts, as they cannot rely on migrants to rebuild spawning biomass and genetic diversity. The northern Caribbean stock itself, as the possibly last remaining healthy stock of queen triggerfish, needs specific attention in order to maintain sufficient genetic diversity for the species to persist in the long term. The results of the present study indicate that genetic diversity is still high but effective population size estimates for recent years suggest that the stock may have experienced a bottleneck and fall below levels necessary for long-term sustainability. Monitoring of effective population size in coming cohorts would be beneficial to ensure that the population maintains enough genetic diversity and develop appropriate protection measures if needed.

This work also led to the development of genomic resources for queen triggerfish. These resources will be beneficial to future analyses of effective population size and adaptive variation in this species that may be important to the conservation of local stocks. Further sequencing is needed to improve the current draft reference genome and enable accurate localization of genetic variants of interest in the queen triggerfish genome.

5. Diffusion of the results and outreach

One short manuscript was written during the course of the project and published in the proceedings of the Gulf and Caribbean Fisheries Institute. The first manuscript reports the testing of microsatellites for genetic studies of queen triggerfish conducted during preliminary phases of the project and summarizes initial results of genotyping by sequencing (before the development of the draft genome).

Two manuscripts, one describing the development of genomic resources and a second one reporting the population genetic survey are in progress and we anticipate submission to peer-reviewed journals.

5.1 Manuscripts

5.1.1 Published

Antoni, L., Cummings, N.J., Saillant, E. 2018. A first Assessment of Genome Wide Genetic Variation and Population Structure in Queen Triggerfish, *Balistes vetula*. Proceedings of the 70th Gulf and Caribbean Fisheries Institute Conference, Merida, Mexico. Gulf and Caribbean Fisheries Institute, pp 262-265.

5.1.2 In preparation

J. Horne, P. Rastas, T. Hildahl†, K. Jones, R. O'Rourke, E.A. Saillant. Linkage mapping and comparative genomics of queen triggerfish *Balistes vetula*.

L. Antoni†, J. Horne, N.J. Cummings, L. Reynal, E. A. Saillant. Population structure of queen triggerfish in the Antilles and South Florida.

5.2 Presentations at conferences

L. Antoni†, L., Cummings, N.J., Saillant, E.A., 2017. A first Assessment of Genome Wide Genetic Variation and Population Structure in Queen Triggerfish, *Balistes vetula*. 70th Gulf and Caribbean Fisheries Institute conference, Merida, Mexico, November 6-10 2017.

† Graduate student

5.3 Student training

One undergraduate student from the Mississippi Gulf Coast Community College (M. Danielle Crawley) was trained to laboratory techniques during the course of the project (2015-2016).

One Ph.D student (Luca Antoni) completed his dissertation research on queen and gray triggerfish conservation genetics (Graduated in Spring 2017).

Another graduate student (Tami Hildahl) received training in laboratory techniques to conduct genome amplifications of small samples and prepare dd-RAD sequencing libraries.

5.4 Other outreach activities

A web page presenting genetic research on gray triggerfish is in development on the PI's laboratory website (<https://sites.google.com/view/gcrl-aquafishgen/>).

A presentation to the public is in the planning for the coming annual workshop organized by the Center for Fisheries Research and Development at the Gulf Coast Research Laboratory (spring 2019). The workshop is attended by fishermen and aims to provide information on research findings as they apply to improve knowledge of fishes and fisheries.

During the course of the project intermediate results obtained and the first manuscript generated were communicated to the NMFS partner (Mrs Cummings) who is closely involved with management of US Caribbean fisheries and contributed to the queen triggerfish assessment conducted in during SEDAR30. The present report and forthcoming manuscripts will be communicated to CFMC, SFMC and GMFMC and project partners in the US Caribbean islands (Puerto Rico DNR, Division of Fish and Wildlife Department of Planning and Natural Resources.).

Outreach in the broader Caribbean was conducted through the presentation given at the 70th annual meeting of the Gulf and Caribbean Fisheries Institute in Merida in November 2017.

6. Data sharing

A bioproject was created in the NCBI SRA database where the raw illumina sequencing reads obtained during genome sequencing and the assembly were deposited. The Bioproject will be

released to the public upon publication of the research, which is expected to occur within one year of the submission of this report.

Horne, J., Saillant, E. 2018. *Balistes vetula* Genome sequencing and assembly. SRA public database, bio project #PRJNA504658 (in construction not released yet)

The Single Nucleotide Polymorphisms discovered and genotyped during the linkage mapping experiment and the population genetic survey will be made available as a variant call format file in the Aquila repository of the University of Southern Mississippi upon acceptance of manuscripts.

6. Acknowledgments

We would like to thank Capt. W. Taylor (blackdog fishing charters, Jupiter, FL), Capt. M. Hanke (light tackle fishing paradise, Fajardo Puerto Rico), Capt. Benigno Rodriguez and the staff of the Fisheries Research Laboratory of the Department of Natural Resources (Puerto Rico), Capt. Daryl Bryan (St Thomas Fishermen Association), Capt. Thomas Daley (St Croix Island), and the staff at IFREMER (La Martinique) for providing samples from La Martinique, and N.J. Cummings and J. Franks for help coordinating sample acquisition. We would also like to thank Dr. Pasi Rastas for help with linkage mapping in Lep-MAP, and Dr. A. Rhine and staff at the New England aquarium for providing queen triggerfish larvae from a spawning event for the linkage mapping experiment. Funding for this work was provided by the NOAA Cooperative Research program project # NA14NMF4540062. Views expressed in this report are those of the authors and do not reflect those of the sponsor.

7. References

- Aiken K.A. 1983 The biology, ecology and bionomics of the triggerfishes, Balistidae. p.191-205. In J.L. Munro (ed.) Caribbean coral reef fishery resources. ICLARM Stud. Rev. 7. 276p.
- Allendorf F.W., P.A. Hohenlohe, G. Luikart 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics*. 11:697–709.

- Andrews K.R., J.M. Good, M.R. Miller, G. Luikart, P.A. Hohenlohe 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet.* 17(2):81-92.
- Angeloni F., N. Wagemaker, P. Vergeer, J. Ouborg 2012. Genomic toolboxes for conservation biologists. *Evolutionary Applications.* 5:130–143.
- Avise J.C. 1998. Conservation genetics in the marine realm. *J. Hered.* 89:377-382.
- Avise J.C. 2010 Perspective: conservation genetics enters the genomics era. *Conservation Genetics* 11:665–669.
- Baird N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko, E.A. Johnson. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3(10): e3376. doi:10.1371/journal.pone.0003376
- Barbato M., P. Orozco-terWengel, M. Tapio, M.W. Bruford. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics* 6:109.
- Begg G.A., K.D. Friedland, J.B. Pearce 1999. Stock identification and its role in stock assessment and Fisheries management: an overview. *Fish. Res.* 43:1-8.
- Bolger A.M., M. Lohse, B. Usadel 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.
- Bourret V., M.P. Kent, C.R. Primmer, A. Vasemägi, S. Karlsson, K. Hindar, P. McGinnity, E. Verspoor, L. Bernatchez, S. Lien 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology* 22:532-551.
- Bradbury I.R., S. Hubert, B. Higgins, T. Borza, S. Bowman, I.G. Paterson, P.V.R. Snelgrove, C.J. Morris, R.S. Gregory, D.C. Hardie, J.A. Hutchings, D.E. Ruzzante, C.T. Taggart, P. Bentzen 2010. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences* 277:3725-3734.
- Carvalho G.R., L. Hauser 1995. Molecular genetics and the stock concept in fisheries. in Carvalho GR, Pitcher TJ (ed) *Molecular genetics in fisheries*. Chapman and Hall, London, pp55–80
- Carson E.W., E. Saillant, M.A. Renshaw, N.J. Cummings, J.R. Gold 2011. Population structure and long-term connectivity and effective size of mutton snapper (*Lutjanus analis*) in the - Caribbean Sea and Florida Keys. *Fishery Bulletin* 109(4):416-428.
- CFMC {Caribbean Fisheries Management Council} 2004. SFA document may be found at: http://www.caribbeanfmc.com/fishery_management_plans.htm.
- Charlesworth B., M. Nordborg, D. Charlesworth 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetic Resources* 70(2): 155-174.
- Chen S. Y. Zhou, Y. Chen, J. Gu 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics*, Volume 34, Issue 17, 1 September 2018, Pages i884–i890,
- Cheviron Z.A., R.T. Brumfield 2009. Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution* 63(6):1593-1605.
- Conover D.O., S.A. Arnott, M.R. Walsh, S.B. Munch 2005. Darwinian fishery science: lessons from the Atlantic silverside. *Canadian Journal of Fisheries and Aquatic Sciences* 62:730-737.

- Cowen R.K., K.M.M. Lwiza, S. Sponaugle, C.B. Paris, D.B. Olson 2000. Connectivity of marine populations: open or closed? *Science* 287:857-859.
- Cowen R.K., C.B. Paris C.B., A. Srinivasan 2006. Scaling of connectivity in marine populations. *Science* 311:522-527..
- Cummings N.J., and Matos-Caraballo 2004. The commercial reef fish fishery in Puerto Rico with emphasis on yellowtail snapper, *Ocyurus chrysurus*: landings, nominal effort, and catch per unit of effort from 1983 through 2003. Sustainable Fisheries Division Contribution (SFD) No. 2004-045 and SEDAR-8 Data Workshop Report (DW) Doc-8. 81p.
- Do C., R.S. Waples, D. Peel, G.M. Macbeth, B.J. Tillett, J.R. Ovenden 2014. NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources* 14:209-214.
- Danecek P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156-2158.
- Dupanloup I., S. Schneider, and L. Excoffier 2002. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11: 2571-2581.
- Durand E., F. Jay, O. E. Gaggiotti, O. François 2009 Spatial Inference of Admixture Proportions and Secondary Contact Zones. *Mol. Biol. Evol.* 26:1963-1973.
- Fox C.W., D.A. Roff, and D.J. Fairbairn 2001. *Evolutionary Ecology: Concepts and Case Studies*. Oxford University Press, UK.
- François O., S. Ancelet, and G. Guillot 2006 Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* 174: 805–816.
- Garrison E., G. Marth 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv. 2012;1207.3907.
- Gasparini J.L. and S.R. Floeter 2001. The shore fishes of Trindade Island, western South Atlantic. *J. Nat. Hist.* 35:1639-1656.
- Glaubitz J.C., T.M. Casstevens, F.Lu, J. Harriman, R.J. Elshire, Q. Sun, E.S. Buckler 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* 9(2): e90346. doi:10.1371/journal.pone.0090346
- Gold J.R., E. Saillant, N.J. Cummings, M.A. Renshaw 2011. Genetic Divergence and Effective Size among Lane Snapper in US Waters of the Western Atlantic Ocean. *North American Journal of Fisheries Management* 31:209-223.
- Grewe P.M., P. Feutry, P.L. Hill, R.M. Gunasekera, K.M. Schaefer, D.G. Itano, D.W. Fuller, S.D. Foster, C.R. Davies. 2015. Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Scientific Reports* 5:16916.
- Harmelin-Vivien, M.L., and J.-C. Quéro 1990. Balistidae. p. 1055-1060. In J.C. Quero, J.C. Hureau, C. Karrer, A. Post and L. Saldanha (eds.) Check-list of the fishes of the eastern tropical Atlantic (CLOFETA). JNICT, Lisbon; SEI, Paris; and UNESCO, Paris. Vol. 2.
- Hare M.P., L. Nunney, M.K. Schwartz, D.E. Ruzzante, M. Burford, R.S. Waples, K. Ruegg, and F. Palstra 2011. Understanding and estimating effective population size for practical application in marine species management. *Conserv. Biol.* 25: 438-449.

- Hedgecock D., A. Pudovkin 2011. Sweepstakes Reproductive Success in Highly Fecund Marine Fish and Shellfish: A Review and Commentary. *Bulletin of Marine Science* 87:971-1002. 10.5343/bms.2010.1051.
- Heyman W.D., S. Kobara, S.J. Pittman, R.S. Nemeth 2013. Caribbean Reef Fish Spawning Aggregations: Biogeography, Future Research, and Management Needs. *Proceedings of the Gulf and Caribbean Fisheries Institute* 66:401-407.
- Hilborn R., Quinn T.P., Schindler D.E., Rogers D.E. 2003. Biocomplexity and fisheries sustainability. *Proc. Natl. Acad. Sci. Am.* 100:6564-6568.
- Hill W.G. 1981. Estimation of effective population size from data on linkage disequilibrium. *Genetics Research* 38:209-216.
- Hinegardner R., D.E. Rosen 1972.. Cellular DNA content and the evolution of teleostean fishes. *American Naturalist* 106: 621-644.
- Hohenlohe P.A., S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson, W.A. Cresko 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6 :e1000862.
- Hollenbeck C.M., Portnoy D.S., Gold J.R. 2016. A method for detecting recent changes in contemporary effective population size from linkage disequilibrium at linked and unlinked loci. *Heredity* 117(4):207-16. doi: 10.1038/hdy.2016.30.
- Horne J.B., I.R. Bradbury, I.G. Paterson, D. Hardie 2016. Complex post-larval dispersal processes in Atlantic cod revealed by age-based genetics and relatedness analysis. *Mar Ecol Prog Ser* 556:237-250.
- Ingram G.W. 2001. Stock structure of gray triggerfish, *Balistes capriscus*, on multiple spatial scales in the Gulf of Mexico. Ph.D Dissertation, Department of Marine Sciences, University of South Alabama.
- Jombart T., S. Devillard, F. Balloux 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94
- Jombart T., I. Ahmed 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–3071
- Laconcha U., M. Iriando, H. Arrizabalaga, C. Manzano, P. Markaide, I. Montes, I. Zarraindia, I. Velado, E. Bilbao, N. Goñi, J. Santiago, A. Domingo, S. Karakulak I. K., Oray, and A. Estonba. 2015. New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*). *PLoS ONE* 10(6):e0128247. doi:10.1371/journal.pone.0128247.
- Lamichhaney S., A. Martinez Barrio, N. Rafati, G. Sundström, C.-J. Rubin, E.R. Gilbert, J. Berglund, A. Wetterbom, L. Laikre, M.T. Webster, M. Grabherr, N. Ryman, L. Andersson. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences USA* 109:19345-19350.
- Larson W.A., L.W. Seeb, M.V. Everett, R.K. Waples, W.D. Templin, J.E. Seeb 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications* 7:355–369.
- Leblois R., A. Estoup, F. Rousset 2003. Influence of Mutational and Sampling Factors on the Estimation of Demographic Parameters in a “Continuous” Population Under Isolation by Distance. *Molecular Biology and Evolution* 20(4):491-502.

- Lee T.N., W.E. Johns, R. Zantopp, and E.R. Fillenbaum. 1996. Moored observations of western boundary current variability and thermohaline circulation 26.5°N in the subtropical North Atlantic. *Journal of Physical Oceanography* 26:962-963.
- Lenormand T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology and Evolution* 17(4):183-189.
- Li C.C., D.E. Weeks, A. Chakravarti 1993. Similarity of DNA finger-prints due to chance and relatedness. *Human Heredity*, 43, 45–52.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv. 2013;1303.3997.
- Limborg M.T., S.J. Helyar, M. De Bruyn, M.I. Taylor, E.E. Nielsen, R. Ogden, G.R. Carvalho, F.P.T. Consortium, D. Bekkevold. 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology* 21:3686-3703.
- Lindeman K.C., R. Pugliese, G.T. Waugh, and J.S. Ault. 2000. Developmental patterns within a multispecies reed fishery: management applications for essential fish habitats and protected areas. *Bulletin of Marine Science* 66(3): 929-956.
- Lischer H.E.L., L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28:298-299.
- Liu J., G. Zapfe, K.-T. Shao, J. L. Leis, K. Matsuura, G. Hardy, M. Liu, and J. Tyler. 2016. *Balistes vetula*. The IUCN Red List of Threatened Species. Version 2016-2. Available: <http://www.iucnredlist.org> (08 November 2016).
- Lotterhos K.E., M.C. Whitlock. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology* 23:2178-2192.
- Luu K, E. Bazin, M.G.B. Blum 2017. pcadapt : an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour* 17:67–77
- Lynch M., K. Ritland 1999. Estimation of pairwise relatedness with molecular markers. *Genetics*,152:1753–1766.
- Manichaikul A., J.C. Mychaleckyj, S.S. Rich, K. Daly, M. Sale, W.M. Chen 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 15;26(22):2867-73.
- Manni F., E. Guérard, and E. Heyer 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by “Monmonier’s algorithm.” *Hum. Biol.* 76:173-190.
- Manooch C.S and C.L. Drennon 1987. Age and growth of yellowtail snapper and queen triggerfish collected from the U S. Virgin Islands and Puerto Rico. *Fisheries Research* 6: 53-68.
- Marcais G., C. Kingsford 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* (2011) 27(6): 764-770
- Mejía, L.S., and A. Acero (Eds.). 2002. Libro rojo de peces marinos de Colombia. INVEMAR, Instituto de Ciencias Naturales - Universidad Nacional de Colombia, Ministerio del Medio Ambiente. La serie Libros Rojos de Especies Amenazadas de Colombia. Bogotá, Colombia. 174 p.
- Nielsen E.E., J. Hemmer-Hansen, N.A. Poulsen, V. Loeschcke, T. Moen, T. Johansen, C. Mittelholzer, G.-L. Taranger, R. Ogden, G.R. Carvalho. 2009. Genomic signatures of local

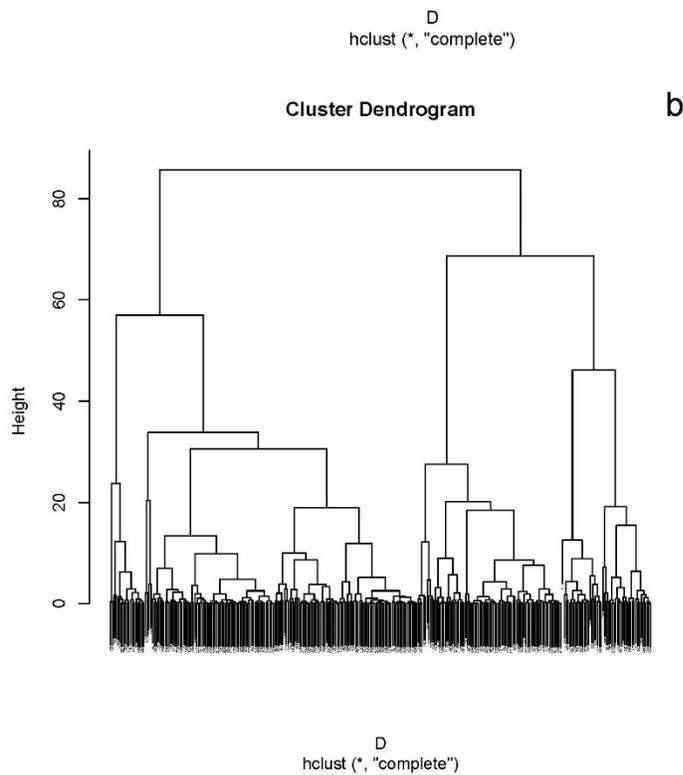
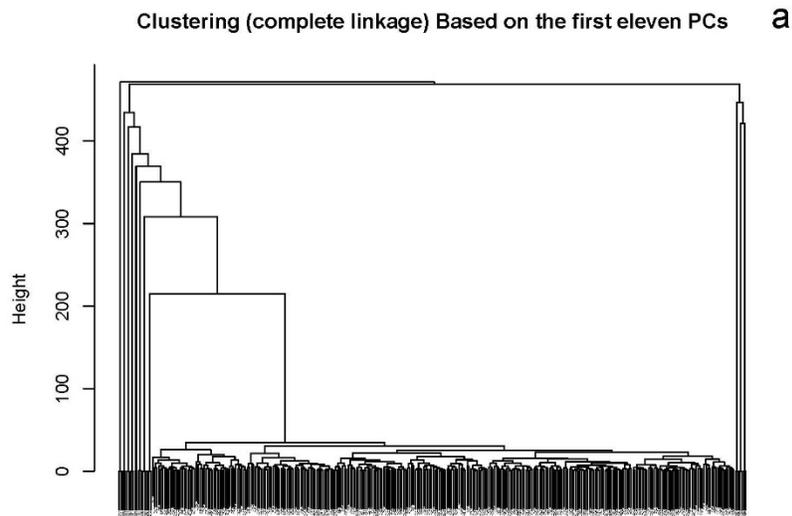
- directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* 9(276):1-11.
- O'Hop J., M. Murphy, and D. Chagaris 2012. The 2012 Stock Assessment Report for Yellowtail Snapper in the South Atlantic and Gulf of Mexico. Assessment report, Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, St Petersburg, FL., 63pp.
- Paradis E., K. Schliep 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2018: 1-3.
- Pembleton L.W., N.O.I. Cogan, J.W. Forster 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* 13(5), 946-952.
- Peterson B.K., J.N. Weber, E.H. Kay, H.S. Fisher, H.F. Hoekstra 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* 7:e37135
- Pew J., P.H. Muir, J. Wang, T.R. Frasier 2015. related: An R package for analysing pairwise relatedness from codominant molecular markers. *Mol Ecol Resour* 15:557-561
- Pritchard J.K., M. Stephens, and P. Donnelly 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Pritchard J.K., X. Wen, D. Falush 2010. Documentation for structure software: Version 2.3. Online manual available at <http://pritch.bsd.uchicago.edu/structure.html>
- Pudovkin A.I., D.V. Zaykin, D. Hedgecock. 1996. On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* 144:383-387.
- Puebla O., E. Bermingham, O. McMillan 2012. On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology* 21:5675-5688.
- Purcell S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, P.C. Sham 2007. PLINK: A Tool Set for Whole-genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81:559-575
- Queller D.C., K.F. Goodnight 1989. Estimating relatedness using molecular markers. *Evolution* 43:258-275.
- Randall J.E. 1967. Food habits of reef fishes in the West Indies. *Stud. Trop. Oceanogr., Univ. Miami* 5:655-847.
- Rastas P. 2017. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data, *Bioinformatics*, 33(23): 3726-3732
<https://doi.org/10.1093/bioinformatics/btx494> .
- Raymond M., F. Rousset 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *The Journal of Heredity* 86:248-249.
- Rieman B.E., F.W. Allendorf 2001. Effective population size and genetic conservation criteria for bull trout. *Journal of North American Fisheries Management* 21:756-764.
- Rios A.B. 2012. A review of the life history characteristics of blue tang and queen triggerfish. SEDAR30-AW-03. SEDAR, North Charleston, SC. 11 pp.
- Rivera J., V. Shervette, N. Pena, K. Flynn, C. Augustine 2014. Queen Triggerfish: Initial Findings on Aspects of Life History for a Data-Poor Species in US Caribbean Waters. Southern Division American Fisheries Society, Charleston, SC January 2014.
- Roberts C. 1996. *Balistes vetula*. In: IUCN 2011. IUCN Red List of Threatened Species. Version 2011.1. <www.iucnredlist.org>.

- Robertson D.R. 1988. Extreme Variation in Settlement of the Caribbean Triggerfish *Balistes vetula* in Panama. *Copeia* 1988:698-703.
- Robins C.R., and G.C. Ray 1986. A field guide to Atlantic coast fishes of North America. Houghton Mifflin Company, Boston, USA. 354 p.
- Rohlf F.J., and G.D. Schnell 1971. An Investigation of the Isolation-By-Distance Model. *Am. Nat.* 105:295-324.
- Rosario A., Kimmel, J.K., and I. Sanchez Ayendez 1986. Survey of commercially exploited fish species and exploratory fishing of underutilized resources around Puerto Rico. Completion report to the National Marine Fisheries Service. Commercial Fisheries Research and Development Act, Program 2-395R, project # P1-88-309. 128pp.
- Rousset F 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219-1228.
- Rousset F. 2008. Genepop'007: a complete reimplement of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8:103-106.
- Ruzzante D.E., Taggart C.T., Cook D. (1999). A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St Lawrence. *Fish. Res.* 43:79-97.
- Sá-Gabriel L.G., W.F. Molina. 2005. Karyotype diversification in fishes of the Balistidae, Diodontidae and Tetraodontidae (Tetraodontiformes). *Caryologia* 58:229-237.
- Saillant E., M.A. Renshaw, N.J. Cummings, J.R. Gold. 2012. Conservation genetics and management of yellowtail snapper (*Ocyurus chrysurus*) in the US Caribbean and South Florida. *Fisheries Management and Ecology* 12:301-312.
- Sambrook J.E., E.F. Fritsch, T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Schweyen H., A. Rozenberg, and F. Leese. 2014. Detection and removal of PCR duplicates in population genomic dd-RAD studies by addition of a degenerate base region (DBR) in sequencing adapters. *Biological Bulletin* 227:146-160.
- Scott W.B., and M.G. Scott 1988 Atlantic fishes of Canada. *Can. Bull. Fish. Aquat. Sci.* 219:731 p.
- SEDAR8 {Southeast Data, Assessment, and Review} 2005. Caribbean yellowtail snapper, *Ocyurus chrysurus*. (a) Stock assessment report, (b) Stock assessment workshop report, and (c) Consensus summary report. SC: SEDAR/NMFS, 1 Southpark Circle #306, Charleston, 21 pp..
- SEDAR-9 (Southeast Data, Assessment, and Review) 2006. Stock Assessment Report of SEDAR-9: Gulf of Mexico Gray Triggerfish. Assessment Report 1. Available: http://www.sefsc.noaa.gov/sedar/Sedar_Workshops.jsp?WorkshopNum=9.
- SEDAR-30 (Southeast Data, Assessment, and Review) 2012. US Caribbean Queen Triggerfish. Final report available at: http://www.sefsc.noaa.gov/sedar/download/S30_Queen_trigger_SAR.pdf?id=DOCUMENT
- Smith M.M., and P.C. Heemstra 1986. Balistidae. p. 876-882. In M.M. Smith and P.C. Heemstra (eds.) *Smiths' sea fishes*. Springer-Verlag, Berlin.
- Stephenson R.L. 1999. Sock complexity in fisheries management: a perspective of emerging issues related to population sub-units. *Fish. Res.* 43:247-249.
- Storz J.F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14: 671-688.

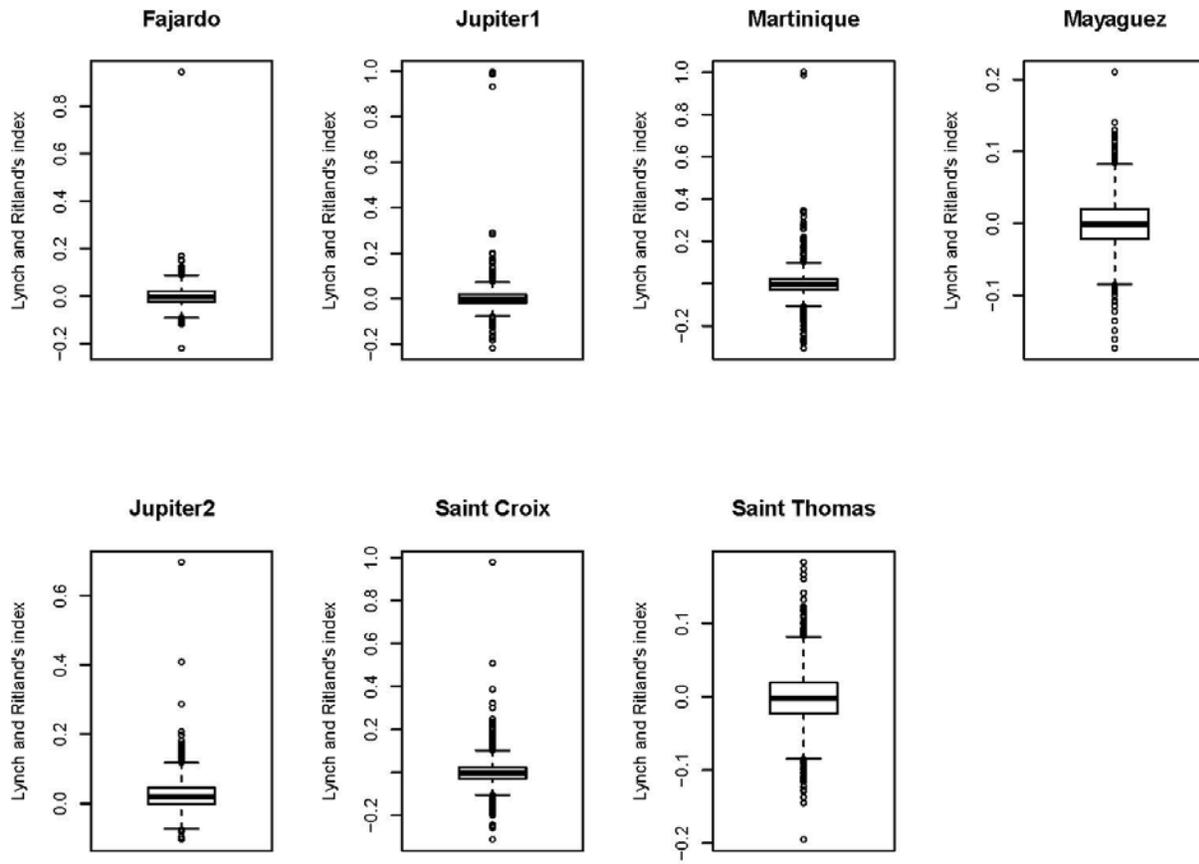
- Swearer S.E., J.E. Caselle, D.W. Lea, and R.R. Warner 1999. Larval retention and recruitment in an island population of a coral-reef fish. *Nature* 402:799-802.
- Wang J. 2002. An estimator for pairwise relatedness using molecular markers. *Genetics*, 160:1203–1215.
- Waples R.S. 1995. Evolutionary significant units and the conservation of biological diversity under the Endangered Species Act. *American Fisheries Symposium* 17:8-27.
- Waples R.S. 1998. Separating the wheat from the chaff: Spatial and temporal patterns of genetic differentiation in marine species. *J. Hered.* 89:438-450.
- 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., C. Do. 2010. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications* 3: 244-262.
- Weir B.S., C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358-1370.
- Whitlock M.C., K.E. Lotterhos 2015. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F_{ST} . *Am Nat* 186:S24–S36.
- Ye C., Z.S. Ma, C.H. Cannon, M. Pop, D.W. Yu 2011. SparseAssembler: de novo assembly with the Sparse de Bruijn Graph. arXiv preprint arXiv:1106.2603.

8. Appendices

Appendix 1. Hierarchical clustering of genotypes for the High_SNP dataset (a) and High_sp_size dataset (b).



Appendix 2. Box plots of the distribution of pairwise relatedness estimates (Lynch and Ritland 1999 estimate) within samples.



Appendix 3. Minimum spanning network of multilocus genotypes based on an Euclidian distance computed from the first two components of principal component analysis.

