A First Assessment of Genome-wide Genetic Variation and Population Structure in Queen Triggerfish, *Balistes vetula*

Evaluación Inicial del Genoma con una Amplia Variación Genética y Estructura de la Población de Pejepuerco, *Balistes vetula*

Une Première Evaluation de la Variation Génétique à l'Echelle du Génome et de la Structure des Populations de Baliste Royal, *Balistes vetula*

LUCA ANTONI[1] NANCIE J. CUMMINGS[2], and ERIC A. SAILLANT[1]*
[1]*School of Ocean Science and Technology — The University of Southern Mississippi, 703 East Beach Drive, Ocean Springs Mississippi 39564 USA.*
*eric.saillant@usm.edu
[2]*NOAA Fisheries Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, Florida 33149 USA.*

ABSTRACT

The queen triggerfish is an economically important Balistid exploited commercially in the Caribbean region for local food markets and the ornamental trade. The species is reported to have declined in several parts of its range yet essential data for its management and conservation such as information on stock structure and demography are still lacking. This work aims to conduct a first assessment of genetic stock structure in the region. Heterologous microsatellite loci developed from gray triggerfish genomic libraries were tested in queen triggerfish. The double-digest Restriction Site Associated DNA (RAD) sequencing method was applied in a survey of genetic variation among six geographic populations including offshore of Jupiter inlet in Southeast Florida, St Croix Island, St Thomas Island, the East and West coasts of Puerto Rico, and La Martinique Island. A total of 3,177 Single Nucleotide Polymorphism (SNP) loci were examined with sample size per locality averaging 71 individuals (range 48 - 85). Allele frequencies were homogeneous among localities with a possible weak isolation-by-distance pattern suggesting long distance dispersal is occurring, a finding consistent with the extended duration of the pelagic larval phase in this species. On-going work focuses on developing reference genomic resources for this species that will be used to strengthen population genetic inferences.

KEY WORDS: Queen Triggerfish, *Balistes vetula*, stock structure, genetics, genotyping by sequencing

INTRODUCTION

The queen triggerfish is a reef fish that frequents tropical and subtropical waters of the Atlantic Ocean. The species is most abundant in the Caribbean region although it is also reported on reef habitats offshore of Bermuda, along the US East coast from Southeastern Florida to the Carolinas, in the Gulf of Mexico, and in South America including Brazil (Robins and Gray 1986). Queen triggerfish have also been observed in the East Atlantic from Cape Verde and the Azores (Harmelin-Vivien and Quéro 1990) to Angola (Smith and Heemstra 1986). In the U.S. Caribbean, the species is exploited by commercial fisheries for local food markets and is also collected alive for the aquarium market. Queen triggerfish are traditionally harvested using traps but the use of hook and line and spear-fishing has developed in recent years in Puerto Rico in parallel with the reduction of the trap fishery in the region (A. Rosario, Laboratorio de Investigaciones Pesqueras, Mayagüez, Puerto Rico, Personal communication).

There are growing concerns regarding the status of Caribbean reef fisheries (Cummings and Matos-Caraballo 2004). While there is a general lack of data on individual species throughout the region, commercial catches, all species combined, have been perceived to decline for several decades in the US Caribbean (Appeldoorn et al. 1992, Matos-Caraballo 2000). A first assessment of queen triggerfish stocks in the U.S. Caribbean was conducted in 2012 (Rios 2012, SEDAR 30). Inferences on the status of the stock during this assessment were very limited due to the lack of essential data on this species. 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. In particular, the complete lack of information on genetic stock structure was identified. Concerns regarding the status of queen triggerfish are not limited to U.S. populations as this species is currently classified as Near Threatened by the International Union for the Conservation of Nature (Liu et al. 2015). This work aims to develop the genomic tools needed for genetic studies of queen triggerfish and provide a first assessment of genetic stock structure in the Caribbean region.

MICROSATELLITE MARKERS

In the first part of this project, 18 microsatellite markers developed from gray triggerfish genomic libraries were evaluated for genetic studies of the queen triggerfish. The characteristics of the developed markers are described in details in Antoni and Saillant (2012).

METHODS

Fin tissue samples from queen triggerfish specimens (n = 32) caught off St. Thomas were provided by Dr. David Olsen. Tissue samples were preserved in a buffered salt-saturated dmso fixative fixative until DNA extraction. DNA was isolated using a phenol-chloroform method (Sambrook et al. 1989). Microsatellites were evaluated in single marker Polymerase Chain Reactions (PCR). PCRs were performed in a total reaction volume of 5 µl containing 6 - 12 ng of genomic DNA, 2

pmol of each primer, 1 nmol of dNTPs, 7.5 nmol of $MgCl_2$, 0.25 U of TAQ polymerase (Promega), and 1X of buffer (Promega). PCR cycles consisted in an initial denaturation at 95°C for 5 min, 35 cycles of 95°C for 30 s, Ta for 30 s, 72°C for 45 s, and a final extension of 15 min at 72°C. Evaluation began at the optimal annealing temperature for amplification of each microsatellite in gray triggerfish but additional temperatures were also evaluated when initial PCR tests were not successful. The 5' end of one of the two primers of each selected primer pair was labeled using a fluorescent dye to allow detection of PCR products on an automated sequencer. The PCR products for each microsatellite were run on a ABI-377XXL sequencer (Applied Biosystems) according to instructions from the manufacturer and evaluated for polymorphism and interpretability of the obtained electropherograms. The size of the PCR products was determined with reference to a size standard prepared as described in DeWoody et al. (2004).

The number of alleles (A), and the observed (Ho) and expected heterozygosity (He) statistics were calculated using the software ARLEQUIN v. 3.01 (Excoffier et al. 2010). Conformance of genotypic proportions to Hardy-Weinberg (H-W) expectations was tested using exact tests in the software GENEPOP v. 4.2 (Raymond and Rousset 1995, Rousset 2008). Probability values were estimated using a Monte Carlo Markov Chain method with 5,000 dememorizations, 500 batches, and 5,000 iterations per batch. The possible occurrence of artifacts impacting scoring including null alleles, large allele dropout and stuttering was tested using Microchecker v. 2.2.3 (Van Oostehout et al. 2004).

RESULTS

Initial amplification tests only led to 8 successfully amplifying loci but re-optimization of annealing temperatures yielded a total of 14 usable microsatellites. The remaining four microsatellites failed to amplify in queen triggerfish (1 locus), did not yield a scorable PCR product (1 locus) or null alleles were found to affect scoring (2 loci). The characteristics of the loci in queen triggerfish (number of alleles, allele size range, heterozygosity, inbreeding coefficients and probability of conformance to Hardy-Weinberg expectations) are summarized in Table 1. While 14 heterologous microsatellite markers tested in this section appear usable in queen triggerfish, assays of individual loci are labor intensive, even if markers can be multiplexed for simultaneous amplification, and the resulting marker density is still low. Further efforts to study queen triggerfish populations considered alternative approaches to genotype samples described below.

RESTRICTION SITE ASSOCIATED DNA SEQUENCING

Recently developed protocols for genotyping based on next generation sequencing technologies such as the double digest Restriction site Associated DNA (dd-RAD) sequencing protocol (Peterson et al. 2012) allow surveying very large numbers (potentially thousands) of Single Nucleotide (SNP) markers for costs and efforts comparable to those involved in genotyping a panel of microsatellites such as the one discussed above. The study of queen triggerfish populations therefore focuses on this approach to genotype and characterize populations.

Table 1. Results of tests of 18 microsatellites developed from gray triggerfish genomic libraries in queen triggerfish (n = 32 samples). Locus are fully described in Antoni and Saillant (2012).

locus	Amplification issues	Scoring artifacts	A	Allele size range	H_e	F_{IS}	P_{H-W}
BC13	allele size shifts	Null alleles	19	231-274	0.925	0.291	0.0026
BC14			3	89-105	0.552	0.155	0.1246
BC16			12	249-287	0.885	0.081	0.0776
BC17			7	212-224	0.757	0.092	0.4032
BC19			5	203-251	0.57	0.068	0.5296
BC2	Not scorable	-	-	-	-	-	-
BC25		Null alleles	9	139-165	0.881	0.539	0
BC26			17	221-277	0.868	0.064	0.0027
BC27			5	151-161	0.772	-0.093	0.2714
BC3			5	132-144	0.259	0.035	0.4805
BC34			17	233-273	0.909	0.038	0.3647
BC36			7	267-279	0.763	0.017	0.5129
BC41			10	136-154	0.87	-0.006	0.9534
BC44			8	239-261	0.497	-0.006	0.4687
BC45			7	242-256	0.591	-0.217	0.4168
BC46			5	103-125	0.584	-0.017	0.5705
BC47			22	112-164	0.914	-0.026	0.8556
BC49	No amplification	-	-	-	-	-	-

Number of alleles (*A*), expected heterozygosity (H_e), inbreeding coefficient (F_{IS}), probability of departure from Hardy-Weinberg equilibrium (P_{HW})

SAMPLING

Sampling focused on US waters of St Thomas (n = 100 collected by trapping), St Croix (n = 100, collected by trapping), Puerto Rico West and East coasts (n = 100, collected by diving and spearfishing off Mayaguez and n = 100, collected by trapping off Fajardo and along the southeast coast respectively), and South Florida (n = 100, collected via hook and line fishing in the Jupiter inlet area). Samples from La Martinique (n = 67) were also provided by IFREMER for analysis.

DNA ISOLATION AND DD-RAD SEQUENCING

DNA was isolated from samples using a phenol-chloroform protocol as above. The dd-RAD sequencing libraries were prepared using protocols derived from those described in Baird et al. (2008) and Peterson et al. (2012). Briefly, following a double digest using SPEI and Sau3AI, RAD-specific Illumina sequencing adapters (Y-yolked to reduce amplification from Sau3AI-Sau3AI fragments, modified from Baird et al. 2008) were ligated to the digested DNA, and amplified to enrich for SPEI-Sau3AI fragments. The resulting dd-RAD libraries were size selected and sequenced (paired-end) on the Illumina Hi-Seq genome sequencing platform.

DNA quality was variable and, for several samples, was too low to perform the dd-RAD-sequencing protocol. We were able to process a total of 470 specimens for library preparation: 79 from Southwest Florida, 79 from East Puerto Rico, 79 from West Puerto Rico, 88 from Saint Thomas, 79 from Saint Croix, and 66 from La Martinique. Sequencing runs included pools of 100 multiplexed libraries except for the initial run that included 75 libraries.

Sequencing yielded an average of 1,502,929 raw reads per samples (1,014,097 after discarding duplicate sequences). After stringent data filtering, 427 individuals were retained in the final dataset (Southwest Florida 78, East Puerto Rico 76, West Puerto Rico 72, Saint Thomas 85, Saint Croix 68, La Martinique 48). The total number of SNPs shared among all localities after filtration was 3,177 (Table 2).

INITIAL DATA ANALYSES AND RESULTS

Preliminary processing of the data revealed very weak levels of divergence among samples ($F_{ST} \leq 0.0007$) and a weak pattern of isolation by distance which is consistent with the life history of the species that involves sedentary behavior at the adult stage but a prolonged pelagic larval phase (63 to 83 days, Robertson 1988 corrected by Lindeman et al. 2000). Estimates of effective population size also suggested a smaller population size in South Florida but were possibly impacted by the lack of information about linkage of markers. Further work focuses on generating genomic resources for queen triggerfish in order to enable accounting for linkage among markers when assessing population structure and potential signatures of natural selection, and also to improve the accuracy of estimates of effective population size by restricting inferences to pairs of unlinked loci.

ACKNOWLEDGMENTS

We wish to acknowledge collaborators involved in sampling efforts, in particular, Capt. William Taylor (Blackdog fishing charters), Capt. Thomas Daley, Capt. Daryl Bryant and Dr. David Olsen (St Thomas Fishermen Association), Capt. Marcos Hanke, Capt. Benigno Rodriguez, Dr. Lionel Reynal and collaborators (Ifremer, La Martinique), and Dr. Aida Rosario and collaborators at the Laboratorio de Investigaciones Pesqueras (Mayaguez, Puerto Rico). This project was supported by the National Oceanic and Atmospheric Administration NOAA Fisheries Cooperative Research Program (#NA14NMF4540062). Views expressed in this paper are those of the authors and do not necessarily reflect those of the sponsors.

LITERATURE CITED

Antoni, L. and E. Saillant. 2012. Development and characterization of Microsatellite markers in the gray triggerfish (*Balistes capriscus*). *Conservation Genetics Resources* **4**:629-631.

Appledoorn, R., J. Beets, J. Bohnsack, B. Bolden, D. Matos, S. Meyers, A. Rosario, Y. Sadovy, and W. Tobias. 1992. Shallow water reef fish stock assessment for the U.S. Caribbean. *NOAA Technical Memorandum NMFS-SEFSC-304*, FL: SE Fisheries Center, Miami, Florida USA. 70 pp.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko, and 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

Table 2. Summary statistics in queen triggerfish sampling locations based on genotypes at 3,177 SNP loci.

	South Florida	East Puerto Rico	West Puerto Rico	St Thomas	St Croix	La Martinique	All
# of individuals	75	76	72	85	68	44	427
# of polymorphic loci	3,157	3,163	3,165	3,161	3,167	3,115	3,177
gene diversity	0.203	0.156	0.161	0.145	0.161	0.170	0.164
gene diversity s.d.	0.096	0.074	0.077	0.069	0.077	0.081	0.078

Summary statistics per population overall loci. s.d.: standard deviation.

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

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

Harmelin-Vivien, M.L. and J.-C. Quéro 1990. Balistidae. Pages 1055-1060 in: J.C. Quero, J.C. Hureau, C. Karrer, A. Post, L. Saldanha, (Eds.). *Check-list of the Fishes of the Eastern Tropical Atlantic (CLOFETA) Volume 2*. JNICT, Lisbon; SEI, Paris; and UNESCO, Paris, France.

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.

Liu, J., G. Zapfe, K.-T. Shao, J.L. Leis, K. Matsuura, G. Hardy, M. Liu, and J. Tyler. 2015. *Balistes vetula* (errata version published in 2016). *The IUCN Red List of Threatened Species* **2015**:e.T2539A97664057. Downloaded on 07 April 2018.

Matos-Caraballo, D. 2000. Overview of Puerto Rico's small scale fisheries statistics: 1994–1997. *Proceedings of the Gulf and Caribbean Fisheries Institute* **51**:215–231.

Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher, and H.E. Hoekstra. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* **7**(5):e37135. doi: 10.1371/journal.pone.0037135

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

Rios, A.B. 2012. A review of the life history characteristics of blue tang and queen triggerfish. *SEDAR30-AW-03*. SEDAR, North Charleston, South Carolina USA. 11 pp.

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, Massachusetts USA. 354 pp.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**:103-106.

Sambrook, J.E., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual, 2nd Edition.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York USA. 1,659 pp.

SEDAR-30 (Southeast Data, Assessment, and Review) 2012. *U.S. 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. Pages 876-882 in: M.M. Smith, P.C. Heemstra, (Eds.) *Smiths' Sea Fishes*. Springer-Verlag, Berlin, Germany.

Van Oosterhout, C.,W. Hutchinson, D. Wills, and P. Shipley. 2004. Micro -checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources* **4**:535-538.