Isolation and characterization of 21 novel microsatellite loci in sailfish *Istiophorus platypterus* Shaw 1792 from a shotgun genomic library

GRISSELMA G. RUBIO-CASTRO¹, ADRIAN MUNGUIA-VEGA^{2,3}, CASIMIRO QUIÑONEZ-VELÁZQUEZ¹, FRANCISCO J. GARCÍA-RODRÍGUEZ^{1*}

¹Instituto Politécnico Nacional–Centro Interdisciplinario de Ciencias Marinas, Departamento de Pesquerías y Biología Marina. Av. Instituto Politécnico Nacional s/n, Col. Playa Palo de Santa Rita 23096, Apdo, Postal 592, La Paz, Baja California Sur, Mexico.

² Conservation Genetics Laboratory, School of Natural Resources and Environment, BSE-317, University of Arizona, 1311 E 4th Street, Tucson, AZ 85721, USA.

³ PANGAS Science Coordination, Comunidad y Biodiversidad A.C. Isla del Peruano 215, Col. Lomas de Miramar, 85448, Guaymas, Sonora, Mexico.

*Corresponding author: Francisco J. García-Rodríguez: e-mail: fjgarcia@ipn.mx Tel. +55 (612) 122-5366, Fax +52 (612) 122-5322

griselma_rubio@hotmail.com, airdrian@email.arizona.edu, cquinone@ipn.mx

Running title: Isolation of microsatellite loci in sailfish

Acknowledgments

This, work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT) of Mexico (Grant SEP-CONACYT-60376) and Secretaría de Investigación y Posgrado-Instituto Politécnico Nacional. GGRC was recipient of a graduate fellowships from the Consejo Nacional de Ciencia y Tecnología, Beca de Estímulo Institucional de Formación de Investigadores, and Comisión de Operación y Fomento de Actividades Académicas-Instituto Politécnico Nacional. CQV and FJGR are fellows of Comisión de Operación y Fomento de Actividades Académicas-Instituto Politécnico Nacional, Estímulos al Desempeño de los Investigadores-Instituto Politécnico Nacional and Sistema Nacional de Investigadores-Consejo Nacional de Ciencia y Tecnología.

Introduction

In the Eastern Pacific Ocean, catch records show high abundance of *Istiophorus platypterus* from the Gulf of California to Ecuador (Kume & Joseph, 1969). Sailfish is an important resource for commercial and recreational fishing in Mexico, and accounts for about 40% of the total catch for billfish, over 50% of the catch for recreational fishing and can reach up to 90% in some ports (Arias-Olaiz, 2007), also represent a considerable economic income for the country. Nevertheless, few studies have been conducted about the structure and genetic diversity of sailfish, employing mitochondrial DNA markers, and microsatellite loci developed for blue marlin *Makaira nigricans* (Buonaccorsi & Graves, 2000; McDowell & Graves, 2002; Lu, Alvarado-Bremer, McKenzie, & Chiang, 2015; Rubio-Castro, Quiñonez-Velázquez & Garcia-Rodríguez, 2016). The amount and spatial scale of genetic structure provides useful information for the management and conservation species, but the development of species-specific molecular markers is essential to avoid potential issues with cross-amplification (Selkoe & Toonen, 2006). Microsatellites (simple sequence repeats, SSR) are one of the most widely used molecular markers for population genetics, fisheries, paternity testing, molecular breeding and conservation (Oliveira, Padua, Zucchi, Vencovsky & Vieira, 2006). Although *I. platypterus* is one of the most important fishery resources, microsatellites for this species had not been described before. In

this study, we isolated and characterized the first set of microsatellite markers for *I. platypterus* from a shotgun genomic library.

Materials and Methods

Total genomic DNA from one specimen from *I. platypterus* obtained from Oaxaca, México was isolated using the DNeasy tissue kit (QIAGEN) of muscle tissue. After RNAase treatment. A shotgun genomic library was constructed using ~5 µg of genomic DNA and sequenced using 454 GS-FLX Titanium chemistry (Roche Applied Science). We searched di-nucleotide microsatellite loci containing at least 10 perfect repeats and tetra-nucleotide loci with at least 8 perfect repeats, and designed primers using the software QDD (Meglecz et al., 2010). We used unique sequence reads larger than 120 base pairs (bp) and consensus sequences within contigs grouping sequences with a similarity $\geq 95\%$ in regions > 30 bp flanking repeats to design primers. We tested primers in 32 individual samples collected in Oaxaca, Mexico. Amplification was performed using the M13-tailed primer protocol for fluorescence labelling of PCR fragments (Schuelke, 2000), where each forward primer was 5' tailed with the M13 sequence. PCRs were conducted in 15 µL volumes with 20-40 ng genomic DNA, 1× Colorless GoTaq Flexi buffer, 0.2 mM of dNTPs, 1.5 mM MgCl₂, 0.3 U of GoTaq DNA polymerase (Promega, Madison, USA), 0.02 µM of the unlabeled M13-tailed forward primer, and 0.2 µM of the fluorescently labeled M13 primer, and 0.2 µM of the reverse primer. A PCR touchdown protocol consisting of 94 °C for 5 min, 15 cycles of 94 °C for 30 s, 65-50 °C for 30 s (1 °C decrease each cycle), 72 °C for 30 s, followed by 40 cycles at 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 5 min was applied. PCR products were checked to confirm expected size on 1% agarose gels. Microsatellites were genotyped using an Applied Biosystems' 3730XL sequencer and alleles were scored using GeneMarker 2.4 software (SoftGenetics, State College, PA). Allele sizes were assigned bins using FLEXIBIN (Amos et al., 2007). Observed and expected heterozygosities and number of alleles were calculated using GENALEX 6.5 (Peakall & Smouse, 2012). We used MICROCHECKER 2.2.3 (Van Oosterhout, Hutchinson, Wills & Shipley, 2004) to test for genotyping errors and presence of null alleles. Deviations from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were estimated using GENEPOP 4.2 (Raymond & Rousset, 1995) and FSTAT 2.9.3.2 (Goudet, 1995), respectively. A sequential Bonferroni test for multiple comparisons was used to adjust p-values ($\alpha = 0.05$).

Results

A total of 118,327 unique reads was generated with an average length of 670 bp after quality filtering ($Q > 20$ using a 10 bp sliding window). We obtained 124 di-nucleotide and 30 tetra-nucleotide microsatellite loci that met our criteria, from which we synthesized and tested 30 tetra-nucleotides and 30 di-nucleotides with the highest number of repeats (> 8). We successfully amplified 12 di-nucleotide and 11 tetra-nucleotide loci in 32 individuals (Table 1), from which two loci were monomorphic (*Ipla23* and *Ipla42*). The mean number of alleles per locus was 8.38 (range 3-20). The observed heterozygosity ranged from 0.41 to 0.94 (mean 0.72) and expected heterozygosity ranged from 0.34 to 0.92 (mean 0.73). No significant LD was found among pairs of loci. Four loci showed significant deviations from HWE (*Ipla36*, *Ipla40*, *Ipla47* and *Ipla57*) after the sequential Bonferroni test ($P < 0.00238$). And, five loci showed null alleles: *Ipla2*, *Ipla40*, *Ipla47*, *Ipla57*, and *Ipla35*.

Discussion

Of the 21 microsatellite loci development using next-generation sequencing, 15 microsatellite loci were in agreement with HWE and did not show null alleles; thus these could be appropriated for population genetic studies. A better explanation about the factors causing departures from HWE should be obtained after to consider more samples (Waples, 2015). Previously, the structure and genetic diversity of sailfish has been

described using mitochondrial DNA and microsatellites developed for blue marlin *M. nigricans* (Buonaccorsi & Graves, 2000; McDowell & Graves 2002; Bangma, 2006; Lu et al., 2015; Rubio-Castro et al., 2016), the polymorphism associated with these 15 novel markers developments in this study suggests their utility for population genetic analysis of sailfish and thus contribute to our understanding of the genetic population structure of the specie in the eastern Pacific. The loci developed here could represent a valuable tool towards the effective fisheries management of populations from *I. platypterus*.

References

Amos, W., Hoffman, J.I., Frodsham, A., Zhang, L., Best, S., & Hill, A.V.S. (2007). Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Note*, 7, 10-14.

Arias-Olaiz, F. (2007). Distribución espacio-temporal de la CPUE del pez vela (*Istiophorus platypterus*) en el Pacífico mexicano. Master of Science Thesis, Centro Interdisciplinario de Ciencias Marinas-Instituto Politécnico Nacional, México, 69 p.

Buonaccorsi, V.P., & Graves, J.E. (2000). Isolation and characterization of novel polymorphic tetra-nucleotide microsatellite markers from the blue marlin, *Makaira nigricans*. *Molecular Ecology*, 9, 820-821.

Goudet, J. (1995). Fstat (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, 86, 485-486.

Kume, S., & Joseph, J. (1969). Size composition and sexual maturity of billfish caught by the Japanese longline fishery in the Pacific Ocean East of 130° W. *Bulletin Far Seas Fisheries Research Laboratory*, 2, 115-162.

Lu, C.P., Alvarado-Bremer, J.R., McKenzie, J.L., & Chiang, W.C. (2015). Analysis of sailfish (*Istiophorus platypterus*) population structure in the North Pacific. *Fisheries Research*, 166, 33-38.

McDowell, J.R., & Graves, J.E. (2002). A genetic perspective on Atlantic sailfish stock structure. International Commission for the Conservation of Atlantic Tunas, Collected Volume of Scientific Papers, 54, 805-810.

Meglecz, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N., & Martin, J.F. (2010). QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*, 26, 403-404.

Oliveira, E.J., Padua, J.G., Zucchi, M.I., Vencovsky, R., & Vieira, M.L.C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29, 294-307.

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

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

Rubio-Castro, G.G., Quiñonez-Velázquez, C., & García-Rodríguez, F.J. (2016). Genetic diversity and population structure of Indo-Pacific sailfish *Istiophorus platypterus* in the eastern Pacific. *Fisheries Science*, 82, 241-249.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18, 233-234.

Selkoe, K.A., & Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, 9, 615-29.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., & Shipley, P. (2004). Microchecker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4, 535-538.

Waples, R.S. (2015). Testing for Hardy-Weinberg proportions: have we lost the plot? *Journal of Heredity*, 106, 1-19.

Table 1. Primer sequences and characteristics of 21 microsatellite loci isolated from *I. platypterus*.

Locus name	Repeat motif	Primer sequence (5'–3')	Size range (bp)	N_A	Ho - H _e	P _{HWE}	GeneBank Accession No.
Ipla2	gcac ₍₁₄₎	F: GGTGATTGGAGCAGTGCAA R: ATCAGCCTCACAGCAGGAAA	123-157	8	0.688- 0.836	0.3724	KY806556
Ipla4	cttt ₍₁₂₎	F: TGAATGTGTTCCACCCACTG R: TTGAATGATGAGGTGGTGTGA	173-189	5	0.625- 0.604	0.5189	KY806557
Ipla6	ctat ₍₁₁₎	F: CCCTGTGATGAAATGTGCTG R: TGCCTGCATGTTAGCTTTTG	333-448	16	0.882- 0.917	0.0186	KY806558
Ipla12	catc ₍₁₀₎	F: GAGAGGGATTTACATTGGAGTCA R: CATGCAACCTGGATAAGTGG	215-255	8	0.719- 0.731	0.2524	KY806559
Ipla17	agac ₍₉₎	F: TGTGGCACTGAGTGTCTGGT R: CTTGCAGCTGCCCTCTGT	179-195	5	0.844- 0.757	0.8058	KY806560
Ipla20	agac ₍₈₎	F: GTGGTTGCCTAGCAACAGGT R: GGACGGGCTCTGTTCTTCTA	146-154	3	0.656- 0.493	1.0000	KY806561
Ipla21	gttt ₍₈₎	F: GTGTGAATGCGTGTGAGACC R: GAGGAGACCTGTGAGTGGGA	167-184	5	0.625- 0.606	0.3786	KY806562
Ipla24	attt ₍₈₎	F: GGGTGGATCTTCACAACGTC R: CTGTTTGCTGCCAGTTGTTT	241-261	3	0.406- 0.344	0.7727	KY806564
Ipla25	acag ₍₈₎	F: CCTCAGAACCACAGACCTGC R: CAGCAGCAACATGAGTCATTC	130-148	4	0.677- 0.583	0.9443	KY806565
Ipla26	tgga ₍₈₎	F: GTAAGCAAGGCTGGATGGAG R: ACCCTGTGCTGCTATATGGG	168-227	7	0.656- 0.697	0.1701	KY806566
Ipla35	ac ₍₃₃₎	F: GCATCATGTGATTGGGTTGT R: TCCAGCAACAATCTGCAAAG	160-184	13	0.780 - 0.883	0.1299	KY806567
Ipla36	ca ₍₃₂₎	F: CTCCTGTTGCTTGGAAGCTC R: TGCCATAGTTAGAAGTCCG	138-157	9	0.750- 0.743	0.0000*	KY806568

131

Ipla40	ac ₍₃₀₎	F: CTGCAACACCTCAGGAAACA R: CCATCACAACTTACAAACGTCA	171-187	6	0.414- 0.675	0.0009*	KY806569
Ipla45	tg ₍₂₈₎	F: GTGGGCACATTTGTCTTGG R: TTCTGTGTTAAGCCCAGGGT	128-173	20	0.935- 0.917	0.0046	KY806571
Ipla47	ca ₍₂₇₎	F: TCCATCACTCGCTCCCTTAC R: CAGTCGTTGCTTCGTTTGTG	166-186	9	0.531- 0.831	0.0000*	KY806572
Ipla51	tg ₍₂₆₎	F: AGCCTGGTCCCAAATAACCT R: GATCAGAACCTGAAACCCGA	370-380	6	0.824- 0.771	0.5465	KY806573
Ipla55	tg ₍₂₅₎	F: ATAAAGGCCTCGACTCTCCC R: CCTGTAAGCTGCCTGTAGCC	237-255	8	0.813- 0.771	0.2289	KY806574
Ipla56	ac ₍₂₅₎	F: ACTACTTCAACACCCGGGGA R: CTCCTCGGTATCCAGCACTT	137-158	10	0.813- 0.801	0.0051	KY806575
Ipla57	tg ₍₂₅₎	F: CATGAGCACTTCAGGGAGAA R: GCTCTGCATCGTCATACCCT	140-175	12	0.719- 0.882	0.0000*	KY806576
Ipla59	tg ₍₂₄₎	F: AACCACCTTTGGGTTGCCATA R: TGTGCAAAGCCTTTTGATTT	236-245	4	0.813- 0.664	0.0350	KY806577
Ipla60	tg ₍₂₄₎	F: GTTAAGGCTCGGTTTAGGGG R: TGGTCCCTACACAGGACACA	176-215	15	0.933- 0.874	0.3723	KY806578

N_A , allele numbers; H_O , observed heterozygosity; H_E , expected heterozygosity; and GeneBank accession number. * Significance of deviation from Hardy–Weinberg equilibrium after Bonferroni correction ($P < 0.00238$).