Characterization by next-generation sequencing of 24 new microsatellite loci for the barred sand-bass, *Paralabrax nebulifer* (Girard, 1854), from the Baja California Peninsula, Mexico

José F. Domínguez-Contreras[1,2], Adrian Munguia-Vega[3,4,], Alejandro Castillo-Lopez[5], José Alberto Gracia-Olea[5], Cecilia M. Blasco[6], and S. Hoyt Peckham[6,7]

[1] Departamento Académico de Biología Marina, Universidad Autónoma de Baja California Sur, La Paz, Baja California Sur, México, 23080.

[2] Instituto Politécnico Nacional, Av. Instituto Politécnico Nacional La Paz, Baja California Sur, México, 23096.

[3] Conservation Genetics Laboratory, School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA, 85721.

[4] PANGAS Science Coordination, Comunidad y Biodiversidad A.C., Guaymas, Sonora, México, 85448.

[5] Pronatura Noroeste, A.C., Ensenada, Baja California, México, 22800.

[6] Smartfish Rescate de Valor A.C., La Paz, Baja California Sur, México, 23000.

[7] Center for Ocean Solutions, Stanford University, Pacific Grove CA USA 93950.

Corresponding author: Adrian Munguia-Vega. E-mail: airdrian@email.arizona.edu

Abstract

We characterized a set of new hypervariable microsatellite loci for the barred sand-bass (*Paralabrax nebulifer*), a marine fish that supports important recreational and artisanal fisheries in California, USA and the west coast of the Baja California Peninsula, Mexico. We performed a shotgun genome sequencing with the 454 XL titanium chemistry and used bioinformatics to search for microsatellite loci with perfect repeats. We selected 40 primer pairs that were synthesized and genotyped in an ABI PRISM 3730XL DNA sequencer in 32 individuals from San Juanico, Baja California Sur. We estimated levels of genetic diversity, deviations from linkage and Hardy-Weinberg equilibrium, the frequency of null alleles and the probability of individual identity for the new markers. We successfully scored 24 microsatellite loci (13 tetranucleotides and 11 dinucleotides). The average number of alleles per locus was 12.5 (range 4 to 23). The average observed and expected heterozygosities were 0.779 (range 0.313 to 0.969) and 0.774 (range 0.350 to 0.939), respectively. We detected significant linkage disequilibrium in two pairs of loci. Genotype frequencies at seven loci showed significant deviations from the expectations of Hardy-Weinberg equilibrium and had estimated null allele frequencies ≥10%. The probability of individual identity for the new loci was 8.5^{-36}. The new markers will be useful for investigating patterns of fine-scale genetic structure and diversity to estimate larval dispersal and assess metapopulation dynamics, information necessary for the sustainable management of *P. nebulifer* fisheries at the west coast of the Baja California Peninsula.

Keywords: Population genetics, Baja California Peninsula, marine connectivity, next generation sequencing, microsatellites, Serranidae

Introduction

The barred sand-bass *Paralabrax nebulifer* (Scorpaeniformes: Serranidae) (Hastings et al. 2014) has been recorded from central California, USA to the Cape Region of Mexico's Baja California Peninsula (including Isla Guadalupe) (Hovey et al. 2002), into the Gulf of California as far north as La Paz Bay and there are also some records from the southern Mexican Pacific coast of Oaxaca (Heemstra 1995).

An important recreational fishery averaging nearly two million fish per year for *Paralabrax nebulifer* that operated in southern California, USA from the 1970s collapsed in 2003 (Erisman et al. 2011) raising concern about the sustainability of other fisheries for *P. nebulifer* through its range. Curiously, on the Pacific coast of the Baja California Peninsula an important artisanal fishery has operated at least since 2001 (Jarvis et al. 2010) with average landings of ~4,000 T/yr over the past decade with an estimated annual worth of ~25 million Mexican pesos (~1.2 millions USD) (CONAPESCA 2016). Despite the recreational and commercial importance of *P. nebulifer* its genetic diversity, population structure and connectivity have not been thoroughly investigated.

Paralabrax nebulifer aggregates to spawn predominantly in July and August in hotspots that have been identified in the waters of south-central California, USA (Jarvis et al. 2014). The aforementioned landings plus spawning aggregations observed by Jarvis et al. (2010) and Jarvis et al. (2014) indicate similar hotspots along the Baja California Peninsula. P. *nebulifer* typically inhabits sandy substrates adjacent to rocky zones at depths between 5-180 m. Allen and Block (2012) calculated a planktonic larval duration of 25.8 ± 2.2 SD days and a settlement size of 9.1 ± 1.1 mm of standard length for *P. nebulifer*, which are critical for the potential of larval dispersal and connectivity among distinct populations (Allen and Block 2012). Connectivity is difficult to measure directly because of many challenges for tracking larvae (Selkoe and Toonen 2011). With the development of genetic tools and analysis methods, particularly highly polymorphic microsatellites markers, it is possible to estimate the degree of larval exchange among populations, mainly through assignment tests (Saenz-Agudelo et al. 2009; Munguia-Vega et al. 2015). Recently, Paterson et al. (2015), using six nuclear microsatellite loci and the mitochondrial control region from *P. nebulifer* found high genetic diversity and a weak but significant genetic structure among locations along California and Baja California Peninsula. A greater number of microsatellite loci could be useful to implement more fine-scale individual-based analyses including estimates of

kinship, paternity, identification of migrant individuals and estimation of recent migration rates, effective population size, and bottlenecks induced by fisheries (Selkoe et al. 2016). In the present study, we tested 40 microsatellite loci specific to *P. nebulifer* in order to study genetic diversity, population structure and connectivity.

Methods

We obtained 32 samples of *Paralabrax nebulifer* from San Juanico, a community that lies in the Gulf of Ulloa on the west coast of Baja California Sur, Mexico. Fin clips were collected from the artisanal fishery in 96% of ethanol and stored at -20 °C. We extracted DNA using the DNeasy blood and tissue kit (QIAGEN, Valencia, CA, USA) following the protocol provided by the manufacturer.

We constructed a genomic library using ~ 5 µg of RNAse-treated genomic DNA from a single individual. The library was sequenced using the 454 XL Titanium chemistry (Roche Applied Science, Indianapolis, IN, USA) at the University of Arizona Genetic Core (UAGC). After eliminating the barcode used to identify the library and applying custom sequence quality criteria ($Q \geq 20$ over a 10 bp window), we obtained 97.56 Mb of sequence distributed in 180,473 individual reads with an average length of 540 bp. We used the iQDD software (Meglécz et al. 2010) to search for perfect di, tri and tetra nucleotide microsatellite loci with at least 10 repeats. We obtained 1,424 loci that met our criteria and for which primers were designed. We tested 40 microsatellite loci (21 di-nucleotides and 19 tetra-nucleotides) with the highest number of repeats. To allow fluorescent labeling, we added the universal M13 primer at the 5´ end of all forward primers (Schuelke 2000). We used four fluorescents dyes (6-FAM, VIC, NED, and PET; Applied Biosystems) and a post-PCR multiplex arrangement (four dyes for each of four different loci sized simultaneously). We performed PCRs in 15µL volumes with 20-40 ng genomic DNA, 1× PCR buffer, 0.2 mM each dNTP, 1.5 mM $MgCl_2$, 0.2% BSA, 0.5 U *Taq* DNA polymerase (Apex, Bioresearch Products), 0.02µM 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. We used 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.

We genotyped PCR products using an Applied Biosystems 3730XL sequencer, and alleles were scored using GENEMAKER Version 2.6.0 (SoftGenetics LLC). Allele sizes were assigned bins using FLEXIBIN (Amos et al. 2007). Observed and expected heterozygosities, number of alleles and the probability of individual identity were calculated using GENALEX 6.5 (Peakall and Smouse 2012). We used MICROCHECKER 2.2.3 (Van Oosterhout et al. 2004) to test for genotyping errors and presence of null alleles. Additionally we calculated the null allele frequency by locus using the Expectation Maximization (EM) algorithm of Dempster et al. (1977) implemented in the software FreeNA (Chapuis and Estoup 2007). Deviations from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were estimated using GENEPOP 4.2 with 10,000 dememorization steps, 1,000 batches and 10,000 iterations (Raymond and 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 and Discussion

From the 40 loci tested, 24 were successfully amplified and genotyped (13 tetra-nucleotides and 11 di-nucleotides) and their sizes ranged from 98 to 310 bp (GenBank Accession numbers KX879728 – KX879751, Table 1). There was significant LD among two pairs of loci ($P < 0.000181$), including *Pane1 – Pane28* and *Pane14 – Pane16*. Seven loci showed significant deviations from HWE (*Pane8*, *Pane16*, *Pane21*, *Pane23*, *Pane28*, *Pane30* and *Pane31*; $P < 0.00208333$) due to a deficit of heterozygote individuals. This deficit could be explained by a Wahlund effect or the sampling of two separate random mating populations, natural selection, assortative mating, inbreeding and/or null alleles (Karlsson and Mork 2005). It is likely that null alleles are the main cause of the deviations from HWE observed, considering these seven loci showed high frequency of null alleles (≥ 18%, except for *Pane31* that showed a null allele frequency of 10%, Table 1). The 17 remaining loci showed no evidence of large allele drop-out, and its null allele frequency was ≤ 13.7% for both EM and Oosterhout methods. The mean number of alleles per locus was 12.52 (range 4 – 23) while the mean number of effective alleles was 9.03 (range 1.54 – 16.38). Observed heterozygosity ranged from 0.313 to 0.969 (mean 0.779) and expected heterozygosity ranged from 0.350 to 0.939 (mean 0.774) (Table 1). The probability that

two individuals will randomly show an identical genotype at the 24 loci was very low (8.5^{-36}). The microsatellite loci developed for *Paralabrax nebulifer* will be useful for on-going research to investigate patterns of larval dispersal, metapopulation dynamics, and fine-scale genetic structure and diversity. In turn, these results will inform the sustainable management of *P. nebulifer* fisheries on the west coast of the Baja California Peninsula.

Acknowledgments

We acknowledge the personnel from the Programa Escama at the Centro Regional de Investigación Pesquera - La Paz and the fishermen who helped us collecting the samples, including Laura Cynthia Zuñiga Pacheco, Ibeth & Yolitzia Bareño Higuera, Francisco Lucero Romero AKA Japo and Rolando Sanchez. Karla Vargas and Stacy L. Sotak helped us at various stages during microsatellite genotyping at the University of Arizona. This work was partially funded by The David and Lucile Packard Foundation grant #2015-62798 supporting the PANGAS Science Coordinator, Fondo Mexicano para la Conservación de la Naturaleza - Fondo Golfo de California grant M-1304-004 to Pronatura Noroeste A.C. and The Walton Family Foundation grant 2016-475 to SmartFish Rescate de Valor, A.C.

References

Allen LG, Block HE (2012) Planktonic larval duration, settlement, and growth rates of the young-of-the-year of two sand basses (*Paralabrax nebulifer* and *P. maculatofasciatus*: fam. Serranidae) from Southern California. Bull S Calif Acad Sci 111:15-21. doi:10.3160/0038-3872-111.1.15

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 7:10-14. doi:10.1111/j.1471-8286.2006.01560.x

CONAPESCA (2016) Comisión Nacional de Acuacultura y Pesca. SAGARPA. http://www.sagarpa.gob.mx/quienesomos/datosabiertos/conapesca/Paginas/default.aspx. Accessed 13 September 2016

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24:621-631. doi:10.1093/molbev/msl191

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B 39:1-38

Erisman BE, Allen LG, Claisse JT, Pondella DJ, Miller EF, Murray JH, Walters C (2011) The illusion of plenty: hyperstability masks collapses in two recreational fisheries that target fish spawning aggregations. Can J Fish Aquat Sci 68:1705-1716. doi:10.1139/f2011-090

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. J Hered 86:485-486

Hastings P, Walker HJ, Galland GR (2014) Fishes: a guide to their diversity. University of California Press. Berkeley.
Heemstra PC (1995) Meros, serranos, guasetas, enjambres, baquetas, indios, loros, gallinas, cabrillas, garropas. In: Fisher W, Krupp F, Scheneider W, Sommer C, Carpenter KE, Niem VH (eds) Guía FAO para la identitifcación de especies para los fines de la pesca. Pacífico centro-oriental, vol III. FAO, Roma, pp 1565-1613
Hovey CB, Allen LG, Hovey TE (2002) The reproductive pattern of barred sand bass (*Paralabrax nebulifer*) from southern California. CalCOFI Rep 43:174-181
Jarvis ET, Linardich C, Valle CF (2010) Spawning-related movements of barred sand bass, *Paralabrax nebulifer*, in Southern California: interpretations from two decades of historical tag and recapture data. Bull S Calif Acad Sci 109:123-143
Jarvis ET, Loke-Smith KA, Evans K, Kloppe RE, Young KA, Valle CF (2014) Reproductive potential and spawning periodicity in barred sand bass (*Paralabrax nebulifer*) from the San Pedro Shelf, southern California. California Fish and Game 100:289-309
Karlsson S, Mork J (2005) Deviation from Hardy–Weinberg equilibrium, and temporal instability in allele frequencies at microsatellite loci in a local population of Atlantic cod. ICES J Mar Sci 62:1588-1596. doi:10.1016/j.icesjms.2005.05.009
Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26:403-404. doi:10.1093/bioinformatics/btp670
Munguia-Vega A, Saenz-Arroyo A, Greenley AP, Espinoza-Montes JA, Palumbi SR, Rossetto M, Micheli F (2015) Marine reserves help preserve genetic diversity after impacts derived from climate variability: Lessons from the pink abalone in Baja California. Global Ecol Cons 4:264-276. doi:10.1016/j.gecco.2015.07.005
Paterson CN, Chabot CL, Robertson JM, Erisman B, Cota-Nieto JJ, Allen LG (2015) The genetic diversity and population structure of barred sand bass, *Paralabrax nebulifer*: A historically important fisheries species off southern and Baja California. CalCOFI Rep 56:97-109
Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539. doi:10.1093/bioinformatics/bts460
Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J Hered 86:248-249
Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. Mol Ecol 18:1765-1776. doi:10.1111/j.1365-294X.2009.04109.x
Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature biotech 18:233-234. doi:10.1038/72708
Selkoe KA, D'Aloia CC, Crandall ED, Iacchei M, Liggins L, Puritz JB, von der Heyden S, Toonen RJ (2016) A decade of seascape genetics: contributions to basic and applied marine connectivity. Mar Ecol Prog Ser 554:1-19. doi:10.3354/meps11792
Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Mar Ecol Prog Ser 436:291-305. doi:10.3354/meps09238
Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535-538. doi:10.1111/j.1471-8286.2004.00684.x

225

Table 1

Table 1. Characteristics of 24 microsatellite loci for *Paralabrax nebullifer*, including: Locus name, GenBank accesion number, Repeat Motif/Dye, forward (F) and reverse (R) primer sequences, expected size in base pairs (bp), size range of observed allelic variation, Number of samples (N), number of alleles (Na), number of effective alleles (Ne), observed (Ho) and expected (He) heterozygosities, probability (*P* value) for deviating from HWE, and estimated null allele frequencies following the Expectation Maximization (EM) and Oosterhout methods. Significant deviations from HWE are shown in bold.

											Null Allele Frequency	
Locus GenBank	**Repeat Motif/dye**	**Primer sequence (5'-3')**	**Expected Size**	**Size Range**	**N**	**Na**	**Ne**	**Ho**	**He**	**HWE**	**EM**	**Oosterhout**
Pane1	$AAAG_{(18)}$ / PET	F: GCAGACTCACATTTTAAGGGG	148	98 - 170	32	17	12.337	0.813	0.919	0.044	0.061	0.0562
KX879728		R: GCACAAACACTTTAAGCAGGG										
Pane4	$AAAG_{(15)}$ / 6-FAM	F: GTCTCTGCCAGCATGTGAAA	272	233 -305	32	18	11.315	0.969	0.912	0.393	0.000	-0.0314
KX879729		R: CCATTGGTGTCGTTCCCTAC										
Pane5	$AAAC_{(14)}$ / NED	F: CCTCCACCAAGGAGGTTATG	240	213 - 253	32	10	6.461	0.844	0.845	0.160	0.011	-0.0035
KX879730		R: CTTGAAATCCCTGCTGCTTC										
Pane6	$AATC_{(14)}$ / VIC	F: CTGTCCGTACTGCATCTCCA	327	277 - 310	32	8	3.938	0.875	0.746	0.498	0.000	-0.0935
KX879731		R: CTCATCATCCAATCCCTGGT										
Pane8	$AGAT_{(13)}$ /NED	F: ATTGTCCCAGGACAGCTCAC	184	157 - 264	31	22	16.288	0.548	0.939	**0.000**	0.203	0.2065
KX879732		R: TATACCCATTCATCCCCACC										
Pane9	$AGAT_{(12)}$ / 6-FAM	F: TGTTTTATTGTCCTACATCATTTCTT	147	124 -194	32	12	3.436	0.813	0.709	0.083	0.011	-0.1004
KX879733		R: CAAAGTCATTTCCCGGTTTG										
Pane11	$AGAT_{(12)}$ / NED	F: ATGCTTGAATTTCCCTCGG	161	113 - 190	32	15	5.069	0.969	0.803	0.003	0.000	-0.1345
KX879734		R: CATCCTCAAGGTCAGCAGTTC										
Pane12	$AGAT_{(11)}$ / 6-FAM	F: CAAAGCACAGTGCGACTCAT	256	233 - 288	32	12	8.904	0.938	0.888	0.531	0.000	-0.0297
KX879735		R: AGGGACGCTCGACTAAAACA										
Pane14	$AGGT_{(10)}$ / VIC	F: AGCAGTTTTCGGGTGATTTG	199	169 - 202	32	8	5.185	0.969	0.807	0.432	0.000	-0.108
KX879736		R: AGTTTGGATAGCGGTGGATG										
Pane15	$ACAG_{(10)}$ / PET	F: ACACAGGGGAGACAGACAGG	152	131 - 173	32	11	3.71	0.844	0.73	0.861	0.000	-0.0868
KX879737		R: CCGTTTATCAACATGTTTACTGTCA										
Pane16	$AGAT_{(10)}$ / PET	F: TGCTCTCCTTGCACTGTCAC	173	130 -230	24	15	8.17	0.25	0.878	**0.000**	0.335	0.3512

KX879738		R: AGAATCAACACTGCCAGAAATG										
Pane18	$ACGC_{(10)}$ / NED	F: AATCAATAGGGCCTGCCTTT	220	204 - 255	32	11	5.52	0.813	0.819	0.363	0.000	0.0101
KX879739		R: CAGCTGTATGTTTTCCATTTGTG										
Pane19	$AGAT_{(10)}$ / 6-FAM	F: TGGTATACCCATTCATCCCC	177	162 - 264	32	22	16.384	0.938	0.939	0.050	0.000	0.0013
KX879740		R: AGGACAGACGGCTCACAAAT										
Pane21	$AC_{(57)}$ / NED	F: GCGTAGGAACTCTGTGGTGAG	251	178 - 254	32	25	18.789	0.594	0.947	**0.000**	0.184	0.1852
KX879741		R: CAGGCTTGCTTGTTTGTCTG										
Pane22	$AC_{(56)}$ / VIC	F: GCGTTATTGTTAAGCCACAGAA	255	160 - 212	32	7	3.531	0.531	0.717	0.007	0.084	0.1377
KX879742		R: TGAGGACTGATTGGACATGC										
Pane23	$AC_{(54)}$ / VIC	F: AATCAGCAGCCTCCATCATC	269	190 - 300	31	27	22.349	0.548	0.955	**0.000**	0.210	0.2117
KX879743		R: TAGGCAGGTGCACTGATGTC										
Pane28	$AC_{(47)}$ / 6-FAM	F: CTGCTGACTGCTGAGTGGAC	194	140 - 231	29	28	20.265	0.414	0.951	**0.000**	0.276	0.2812
KX879744		R: AGCTACTCTCCTGCGCCATA										
Pane30	$AC_{(46)}$ / PET	F: CGTGAGTCAAACATGCTGCT	150	245 - 250	32	3	1.857	0.156	0.461	**0.000**	0.211	0.2851
KX879745		R: AAGCAACAGCAAAACGTGAG										
Pane31	$AC_{(45)}$ / VIC	F: TGAACCACGTCCAGACTGAG	245	190 - 278	29	31	20.512	0.759	0.951	**0.000**	0.098	0.1018
KX879746		R: TTCTGCTCACTTCCAGCATTT										
Pane33	$AC_{(45)}$ / 6-FAM	F: TAGGTGGGGCAGTTTAGACG	292	194 - 206	32	4	1.539	0.313	0.35	0.132	0.042	0.0553
KX879747		R: AAGGGGTGAAGTCCCTGATT										
Pane37	$AC_{(42)}$ / 6-FAM	F: TGACTCACGTTCTGGCTGTT	219	159 - 203	32	13	6.132	0.688	0.837	0.037	0.076	0.0899
KX879748		R: CACAGCTAGTTGGGAAGTTTAGC										
Pane38	$AC_{(41)}$ / PET	F: AGTAGCCTACCGGCCACAG	289	109 - 242	32	9	3.17	0.688	0.685	0.555	0.000	0.0034
KX879749		R: GATGTCCAGGATGGCTTTGT										
Pane39	$AC_{(41)}$ / VIC	F: CCCTCACTGACTGCAGGTG	154	133 - 211	28	13	2.23	0.464	0.552	0.099	0.039	0.0958
KX879750		R: ACGCACCTGATCTGCTTGAT										
Pane40	$AC_{(40)}$ / NED	F: ATAGGCCTGCTGTCTGCACT	196	123 - 205	32	23	9.846	0.781	0.898	0.020	0.058	0.0657
KX879751		R: TGAATGAACATCTGATAAAGATCG										