Mol Biol Rep
DOI 10.1007/s11033-017-4105-4

SHORT COMMUNICATION

Characterization of 32 microsatellite loci for the Pacific red snapper, *Lutjanus peru*, through next generation sequencing

David A. Paz-García[1,2] · Adrián Munguía-Vega[3,4,5] · Tomas Plomozo-Lugo[6] · Amy Hudson Weaver[6]

Received: 7 May 2016 / Accepted: 24 April 2017

Abstract We developed a set of hypervariable microsatellite markers for the Pacific red snapper (*Lutjanus peru*), an economically important marine fish for small-scale fisheries in the west coast of Mexico. We performed shotgun genome sequencing with the 454 XL titanium chemistry and used bioinformatic tools to search for perfect microsatellite loci. We selected 66 primer pairs that were synthesized and genotyped in an ABI PRISM 3730XL DNA sequencer in 32 individuals from the Gulf of California. We estimated levels of genetic diversity, deviations from linkage and Hardy–Weinberg equilibrium, estimated the frequency of null alleles and the probability of individual identity for the new markers. We reanalyzed 16 loci in 16 individuals to estimate genotyping error rates. Eighteen loci failed to amplify, 16 loci were discarded due to unspecific amplifications and 32 loci (14 tetranucleotide and 18 dinucleotide) were successfully scored. The average number of alleles per locus was 21 (±6.87, SD) and ranged from 8 to 34. The average observed and expected heterozygosities were 0.787 (±0.144 SD, range 0.250–0.935) and 0.909 (±0.122 SD, range 0.381–0.965), respectively. No significant linkage was detected. Eight loci showed deviations from Hardy–Weinberg equilibrium, and from these, four loci showed moderate null allele frequencies (0.104–0.220). The probability of individual identity for the new loci was 1.46^{-62}. Genotyping error rates averaged 9.58%. The new markers will be useful to investigate patterns of larval dispersal, metapopulation dynamics, fine-scale genetic structure and diversity aimed to inform the implementation of spatially explicit fisheries management strategies in the Gulf of California.

Keywords Population genetics · Gulf of California · Marine connectivity · Pyrosequencing · Microsatellites · Lutjanidae

✉ Adrián Munguía-Vega
airdrian@email.arizona.edu

[1] Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA

[2] Laboratorio de Necton y Ecología de arrecifes. Centro de Investigaciones Biológicas del Noroeste, Instituto Politécnico Nacional, 195, Col. Playa Palo de Santa Rita Sur, La Paz, Baja California Sur, Mexico

[3] PANGAS Science Coordination, Comunidad y Biodiversidad A.C., Guaymas, Sonora, Mexico

[4] Conservation Genetics Laboratory, School of Natural Resources and the Environment, The University of Arizona, Tucson, AZ, USA

[5] @Genetics Lab., La Paz, Baja California Sur, Mexico

[6] Sociedad de Historia Natural Niparaja A.C., La Paz, Baja California Sur, Mexico

Introduction

The Pacific red snapper (*Lutjanus peru*, Nichols and Murphy, 1922) is an economically important marine fish for artisanal fishers in the west coast of Mexico [1, 2]. Pacific red snapper is a demersal species abundant near the coast in waters <90 m deep, associated to rock and coral reefs, and distributed from California US to Peru [3]. In 2010, 4578 tons of *L. peru* were landed in the western coast of Mexico, from which about half (2209 tons) originated from the Gulf of California [4]. The fishery is considered at its maximum levels of sustainable exploitation [4].

The elucidation of metapopulation dynamics via larval dispersal driven by variable oceanographic currents among localities and seasons are critical for the implementation of spatially explicit management strategies including

Published online: 27 April 2017

marine reserves [5]. In recent years molecular data has been increasingly used to guide marine spatial planning taking into account levels of genetic diversity and structure within species and used as a proxy to estimate the strength and direction of connectivity among populations, including self-recruitment [6]. Despite of the fishery importance of *L. peru* few genetic resources are currently available for the species [3, 7].

Recent advances in next generation sequencing have rendered the discovery of novel microsatellite loci that due to their higher mutation rates are effective tools for studying demographic connectivity and detect geographic structure in marine fishes [8]. Here, we report the first 32 microsatellite loci isolated de-novo for *L. peru*.

Materials and methods

Sampling, genomic DNA extraction, and primer design

We used one fin clip from *L. peru* to extract genomic DNA with the DNeasy Blood and Tissue Kit (QIAGEN). Five micrograms of genomic DNA were treated with RNAse and used to construct a shotgun genomic library that was sequenced at the University of Arizona Genetics Core (UAGC) using the Titanium XL + 454 pyrosequencing chemistry (Roche Applied Science). 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 44.78 Mb of sequence distributed in 82,276 individual reads with an average length of 534 bp. We used the software iQDD [9] to search for perfect di, tri and tetra nucleotide microsatellite loci with at least ten repeats. We obtained 539 loci that meet our criteria and for which primers were designed. To allow fluorescent labeling, the universal M13 primer was added at the 5′ end of all forward primers [10].

PCR conditions and primers selection

We sampled 32 individuals collected in 2015 near San Diego Island (Cabeza de San Diego) located in the southwest Gulf of California (25.18912472 N, 110.7100591 W) with the help of small-scale fishers under permit issued by Comisión Nacional de Pesca # 103053993271-5. Samples were stored in 95% ethanol and kept at −20 °C in the laboratory. Genomic DNA was extracted as described above. For the validation of designed primers, 66 primer pairs were synthesized and tested initially (34 tetranucleotides and 32 dinucleotides) in seven individual samples. PCRs were conducted in 15 µl volumes with 20–40 ng genomic DNA, 1× PCR buffer, 0.2 mM each dNTP, 1.5 mM $MgCl_2$, 0.5 U Taq DNA polymerase (Invitrogen), 0.2 µM of the fluorescently labeled M13 primer, 0.02 µM of the unlabeled M13-tailed forward 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. PCR products were visualized for amplification in 1.5% agarose gels.

From the 66 tested primers pairs, 48 showed successful amplification and were genotyped in 32 individual samples in an ABI PRISM 3730XL DNA sequencer (Applied Biosystems) using GeneScan™ 500 LIZ™ as size standard at the UAGC. Sixteen loci were discarded due to the presence of multiple, unspecific amplifications, while the other 32 loci (14 tetranucleotide and 18 dinucleotide) were successfully scored. To test the reproducibility of the microsatellite loci and the observed genotypes, a subset of 16 samples for 16 loci were amplified a second time as described above and analyzed in order to determine the genotyping error rates (see below).

Data analysis

Raw allele sizes were scored with the software Peak Scanner V.2 (Applied Biosystems) and binned using FLEXIBIN [11]. Number of alleles (Na), Observed (Ho) and expected (He) heterozygosity, the inbreeding coefficient F_{IS} [12], linkage disequilibrium and deviation from the Hardy–Weinberg equilibrium (HWE) were evaluated by locus using GENEPOP 4.5.1 [13] with 10,000 dememorization steps, 1000 batches and 10,000 iterations. A Fisher's exact test was applied with sequential Bonferroni correction ($\alpha = 0.05$) [14]. Null allele frequencies (NaF) and scoring errors caused by stutter peaks or large allele dropout were calculated with Micro-Checker [15]. GENALEX 6.5 [16] was used to calculate the probability of individual identity for the 32 loci described. Genotyping error rates were estimated for 16 microsatellite loci by comparing genotypes between PCR repetitions. Mean error rate per locus was estimated according to [17], using 16 individual samples per locus, and in total 256 genotypes were analyzed a second time for this purpose.

Results and discussion

The 32 loci that were successfully amplified and genotyped in *L. peru* were polymorphic (Table 1). Sequences were deposited in GenBank under accession numbers KU951491-KU951522. The average number of alleles per locus was 21 (±6.87, SD) and ranged from 8 to 34. The average Ho was 0.787 (±0.144 SD, range 0.250–0.935) and He was 0.909 (±0.122 SD, range

Table 1 Characterization of 32 microsatellite loci for the Pacific red snapper (*Lutjanus peru*)

	Locus	Repeat motif	Primer sequence (5′–3′)	Dye label	Allelic size range (bp)	N	Na	Ho	He	F_{IS}	HWE	NaF	GER	GenBank
1	*Lupe01*	GAGT(38)	F:AACTTTGCAGAACCGAGGAG R:TTTGAGATTAGCTGCGGACA	NED	108–285	32	32	0.844	0.965	0.128	*	0.054	0.250	KU951491
2	*Lupe02*	CATT(31)	F:TTCTGTGTCAGCGCCATTAC R:TTTGGACTCAAAGAAGGACGA	FAM	215–409	31	27	0.839	0.960	0.128	NS	0.054	0.063	KU951492
3	*Lupe03*	TCCA(22)	F:CCCAGTTTCTCTTGGGACAA R:CACTGTGACCCAACCTCAGA	FAM	191–285	31	21	0.677	0.948	0.289	**	0.132	–	KU951493
4	*Lupe13*	AGAT(17)	F:TGGCTCCTAAAATGCAGATG R:GGCAGCATTTTATGTCTGTTTG	FAM	199–264	32	14	0.781	0.906	0.139	NS	0.058	0.125	KU951494
5	*Lupe16*	TCTA(15)	F:TGCCCATCTAATGCATTGTAAC R:ATTTCAGCTGTGGGAGCTTT	FAM	234–334	32	21	0.844	0.935	0.099	*	–	0.125	KU951495
6	*Lupe19*	TCTA(15)	F:TTCCTGCAGCTTTAACACCA R:TGAAGGAGGGTAGGTTTCCA	FAM	110–185	32	17	0.719	0.934	0.233	*	0.104	–	KU951496
7	*Lupe21*	TTTA(13)	F:TGATGTCTCTTTAGGAGATTGGG R:ATGAGGGTTGGGCACTTTTT	VIC	110–155	32	11	0.813	0.814	0.003	NS	–	0.063	KU951497
8	*Lupe23*	TGGA(13)	F:GCTCAACATGAAGGGCTGAT R:GCCTGTGACCCCATATAACC	FAM	234–381	32	24	0.906	0.954	0.051	NS	–	0.067	KU951498
9	*Lupe24*	AGAT(13)	F:AGCAACTTTGGGCAGAAGAA R:CATTAATGTGAACTTGTGCAATC	FAM	217–270	30	14	0.767	0.884	0.135	NS	–	0.077	KU951499
10	*Lupe25*	TAGA(13)	F:GCCTATTCCAGCGTGACATT R:ATTTCCCAGTTTGGGATCAA	VIC	260–348	32	13	0.813	0.902	0.101	NS	–	0.000	KU951500
11	*Lupe28*	TGGA(11)	F:TGTAATCCTCATTCTGGCTGG R:TCCACCTCAACACCTTGACA	FAM	98–178	31	18	0.903	0.941	0.041	NS	–	0.250	KU951501
12	*Lupe29*	AGAT(11)	F:CGGACATTCATAATAGAACAACAGA R:CTGCAGTGAGCTGAGCTTTT	FAM	112–160	32	13	0.781	0.908	0.141	NS	0.059	0.000	KU951502
13	*Lupe33*	TGGA(10)	F:ATTTGCTGTGGCTGCTCTCT R:GGTTAGCAATGGAAGCCATC	VIC	153–232	32	18	0.813	0.937	0.135	NS	0.057	0.200	KU951503
14	*Lupe34*	AGAT(10)	F:CTGACTTTCACCTCATGACAGA R:GTTAGGGTAAGGGAGGGCAG	FAM	128–195	30	14	0.900	0.869	−0.036	NS	–	0.063	KU951504
15	*Lupe37*	AC(41)	F:ACCCGGGGTGTTACACATTA R:TCCGTGTCCATGTCCAATTA	FAM	81–161	32	24	0.781	0.951	0.181	NS	0.080	–	KU951505
16	*Lupe38*	AC(40)	F:CTTCGCACCCTCCAATTTTA R:AAAATCTCCAGTGGAAGGAAGA	VIC	160–285	25	26	0.520	0.969	0.469	**	0.220	–	KU951506
17	*Lupe39*	AC(39)	F:CCTTTCATCAGAGCAGAGGC R:CATGTGCACGTTCACTCTCC	VIC	203–269	32	24	0.875	0.955	0.085	NS	–	0.000	KU951507
18	*Lupe42*	AC(38)	F:CACATTCACATCCACTTGGG R:GTGCCTGCAAAGCATATTCA	FAM	116–173	32	22	0.813	0.955	0.152	NS	0.065	–	KU951508
19	*Lupe43*	GT(36)	F:CCAAGATTATTTTGTGTCTGTCTGA R:TTCAACATAAACCTTGCTCCTC	VIC	113–211	29	26	0.793	0.966	0.181	NS	0.079	–	KU951509
20	*Lupe44*	GT(35)	F:CCGTGACTAACACATGGTGC R:CTGCAATGCACTTGTCAGAT	FAM	134–224	32	32	0.875	0.976	0.105	NS	0.043	–	KU951510

Table 1 (continued)

	Locus	Repeat motif	Primer sequence (5′–3′)	Dye label	Allelic size range (bp)	N	Na	Ho	He	F_{IS}	HWE	NaF	GER	GenBank
21	*Lupe45*	GT(35)	F:CACTCTGCAGTTGACCATCG R:ACGGAGTACATGCTGCCTCT	VIC	103–144	32	8	0.250	0.381	0.348	NS	0.091	–	KU951511
22	*Lupe51*	AC(35)	F:GCGTCTTGGTTGGTCATCTT R:CCTATTGACAGTCCGTCCGT	VIC	217–290	31	27	0.935	0.965	0.031	NS	–	–	KU951512
23	*Lupe52*	AC(34)	F:CATGGTGGACTCATGTCTGC R:TTCATGACTCAGAGCCCAAA	FAM	162–245	32	34	0.844	0.967	0.129	*	0.055	–	KU951513
24	*Lupe53*	GT(33)	F:TGGCAAGTATGTAAGCAACAGG R:AGCCTCCAGACTGCAGAGAG	FAM	122–206	31	29	0.677	0.969	0.305	**	0.141	–	KU951514
25	*Lupe54*	GT(32)	F:TGTAGGTGCACATTACAAATATAGACA R:CGTTGTCGAGGAACATCTGA	FAM	133–197	32	25	0.813	0.962	0.157	*	0.069	–	KU951515
26	*Lupe55*	GT(33)	F:GACTGTACCATGTCGAGGGG R:TTGTGCAGGATACGTGCTGT	FAM	234–315	31	27	0.871	0.956	0.090	NS	–	0.000	KU951516
27	*Lupe58*	AC(32)	F:AGCAGCCAATCAACTTCACC R:TTCCAGGTGAAACCTTTGGA	VIC	147–182	32	18	0.813	0.942	0.140	NS	0.059	–	KU951517
28	*Lupe61*	AC(32)	F:TTTGCATGCCAAGAACCAT R:TGCAGGCTGAAATTGAACTG	VIC	190–243	29	24	0.897	0.957	0.064	NS	–	–	KU951518
29	*Lupe62*	GT(31)	F:TGGGAATTAGTGACTGTGAGCA R:GAGCACGACAGCATTAGCAG	VIC	214–288	32	23	0.906	0.941	0.038	NS	–	0.250	KU951519
30	*Lupe63*	AC(34)	F:CAAGCAAATGGAGCAAGACA R:CATTATTGCTCTGATCCCGC	VIC	206–288	32	24	0.906	0.950	0.047	NS	–	0.000	KU951520
31	*Lupe64*	AC(31)	F:TCCACCAAGGCTCAGGTTAC R:GGGTTTCCTCTGTTTGGACA	VIC	163–207	32	18	0.781	0.902	0.136	NS	0.056	–	KU951521
32	*Lupe65*	AC(31)	F:CTGACATCTGTGCTGGCAGT R:CCAGATGCTGTTGCATCAAT	VIC	216–309	29	8	0.448	0.558	0.199	NS	–	–	KU951522

Forward (F) and Reverse (R) primers, sample size (N), number of alleles per locus (Na), observed (Ho) and expected (He) heterozygosity, the inbreeding coefficient (F_{IS})

Null allele frequencies (NaF, Brookfield method). Genotyping error rate (GER) and GenBank accession number (GenBank)

Deviation from Hardy–Weinberg equilibrium (HWE): not significant (NS), *$P < 0.05$, **$P < 0.01$

0.381–0.965). Average inbreeding coefficient (F_{IS}) value was 0.139 (±0.103 SD) and ranged from 0.031 to 0.348 (Table 1). All loci pairs were tested for linkage disequilibrium, but no significant patterns were detected. Fisher's exact tests revealed that 8 out of 32 loci exhibited significant deviations from HWE ($P < 0.05$, Table 1). Micro-Checker detected the probable presence of null alleles at low frequency in 14 microsatellite loci (≤0.091, Table 1). Four microsatellite loci that deviated from HWE (*Lupe03*, *Lupe19*, *Lupe38*, and *Lupe53*) showed moderate null allele frequencies (0.104–0.220, Table 1). Selection, incomplete sampling, mating system effects or fluctuations in population size may also explain deviations from HWE [18]. Genotyping error rates ranged from 0 to <13% in most of the 16 loci reanalyzed (average 9.58%), while only four loci (*Lupe01*, *Lupe28*, *Lupe33* and *Lupe62*) showed error rates between 20 and 25% (Table 1). Allelic dropout caused genotyping errors in these loci where true heterozygous individuals were scored as homozygous in the second run of analyses due to the failure to detect the amplification of one of the alleles in some samples. Error rate in these four loci may underestimate the population diversity in *L. peru*. Nevertheless, a few loci were extremely polymorphic (e.g. *Lupe01* and *Lupe44*) and showed 32 alleles in 32 genotyped individuals, suggesting that genotyping a larger number of individuals would improve the estimates of allelic frequencies. The probability that two individuals will randomly show an identical genotype at the 32 loci was extremely low (1.46^{-62}). The microsatellite loci developed for *L. peru* will be useful for on-going research to investigate patterns of larval dispersal, metapopulation dynamics, fine-scale genetic structure and diversity aimed to inform the implementation of spatially explicit fisheries management strategies in the Gulf of California.

Acknowledgements We would like to acknowledge Mariana Walther, Juan Leonardo Lucero Cuevas (Tito) and Jose Amador Gutierrez (Pepe) for their assistance with acquiring samples in the field, and Ollin Gonzalez-Cuellar for helping with logistics during the project. Alexander Ochoa, Karla Vargas, Jose Francisco Dominguez-Contreras (Borre), Geraldine Parra and Stacy L. Sotak help us at various stages during microsatellite genotyping. DAPG received a CONACYT fellowship (250126). This work was funded by The Walton Family Foundation grant # 2011-1235 and The David and Lucile Packard Foundation grants #2013-39359, #2013-39400, #2015-62798. We also thank two anonymous reviewers for helpful comments that improved this manuscript.

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflict of interest exists.

References

1. Diaz-Uribe JG, Chavez EA, Elorduy-Garay JF (2004) Assessment of the Pacific red snapper (*Lutjanus peru*) fishery in the southwestern Gulf of California. Ciencias Marinas 30(4):561–574
2. Leslie H, Basurto X, Nenadovic M et al (2015) Operationalizing the social-ecological systems framework to assess sustainability. Proc Natl Acad Sci USA 112:5979–5984
3. Rocha-Olivares A, Sandoval-Castillo J (2003) Mitochondrial diversity and genetic structure in allopatric populations of the Pacific red snapper *Lutjanus peru*. Ciencias Marinas 29(2):197–209
4. Zarate-Becerra ME, Espino-Barr E, Garcia-Boa A et al (2014) Huachinango del Pacífico Centro-Sur, costa de Nayarit a Chiapas. In: Belendez-Moreno FJ, Espino-Barr E, Galindo-Cortes G, Gaspar-Dillanes MT, Huidobro-Campos L, Morales-Bojorquez E (eds) Sustentabilidad y Pesca Responsable en Mexico. Evaluacion y Manejo. SAGARPA Instituto Nacional de Pesca, Mexico, D.F., pp 141–175
5. Brown CJ, White C, Beger M, Grantham HS, Halpern BS, Klein CJ, Mumby PJ, Tulloch VJD, Ruckelshaus M, Possingham HP (2015) Fisheries and biodiversity benefits of using static versus dynamic models for designing marine reserve networks. Ecosphere 6(10):1–14. doi:10.1890/es14-00429.1
6. Beger M, Selkoe KA, Treml EA, Barber PH, von der Heyden S, Crandall ED, Toonen RJ, Riginos C (2014) Evolving coral reef conservation with genetic information. Bull Mar Sci 90(1):159–185. doi:10.5343/bms.2012.1106
7. Bayona-Vasquez NJ, Hernandez-Alvarez CA, Glenn T, Dominguez-Dominguez O, Uribe-Alcocer M, Diaz-Jaimes P (2015) Complete mitogenome sequences of the pacific red snapper (*Lutjanus peru*) and the spotted rose snapper (*Lutjanus gutattus*). Mitochondrial DNA 28(2):1–2. doi:10.3109/19401736.2015.1115851
8. Jackson AM, Munguía-Vega A, Beldade R, Erisman BE, Bernardi G (2015) Incorporating historical and ecological genetic data for leopard grouper (*Mycteroperca rosacea*) into marine reserve design in the Gulf of California. Conserv Genet 16:811–822. doi:10.1007/s10592-015-0702-8
9. Meglécz 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(3):403–404. doi:10.1093/bioinformatics/btp67010.1111/j.17550998.2009.02750.x
10. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234
11. 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(1):10–14. doi:10.1111/j.1471-8286.2006.01560.x
12. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evol Int J Org Evol 38:1358–1370
13. Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population Genetics Software for exact tests and ecumenicism. J Hered 86:248–249
14. Rice WR (1989) Analyzing tables of statistical tests. Evol Int J Org Evol 43(1):223–225
15. 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(3):535–538. doi:10.1111/j.1471-8286.2004.00684.x
16. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an

update. Bioinformatics 28(19):2537–2539. doi:10.1093/bioinformatics/bts460
17. Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859. doi:10.1038/nrg1707
18. Allendorf F, Luikart G (2007) Conservation and the genetics of populations. Blackwell, Padstow, UK