CICIMAR OCEÁNIDES

CICIMAR Oceánides
ISSN 2448-9123
https://oceanides.ipn.mx
Centro Interdisciplinario de Ciencias Marinas
https://doi.org/10.37543/oceanides.v31i1.162

Vol. 31 No. 1
Enero – Junio
2016

ISOLATION AND CHARACTERIZATION OF NEW MICROSATELLITE MARKERS FOR THE CORTÉS GEODUCK (*Panopea globosa*)

Bisbal-Pardo, Celia Isabel [1], Miguel Ángel Del Río-Portilla [2], Ana Yonori Castillo-Paéz [1] & Axayácatl Rocha-Olivares [1, *]

[1] Laboratorio de Ecología Molecular, Departamento de Oceanografía Biológica, CICESE, Carretera Ensenada-Tijuana 3918, Zona Playitas, Ensenada, Baja California, México. [2] Departamento de Acuicultura, CICESE, Carretera Ensenada-Tijuana 3918, Zona Playitas, Ensenada, Baja California, México. * Corresponding author. e-mail: arocha@cicese.mx

ABSTRACT. The geoduck *Panopea globosa* is a long-lived and large endemic infaunal clam sustaining a growing fishery in the Northwest coast of México that, in spite of its increasing demand in Asian markets very little is known about its biology. In order to provide genetic markers to support genetic research of wild populations, nine novel microsatellite loci (di-, tri-, and tetranucleotide repeats) were developed using shotgun sequencing with next generation technology (Illumina). The number of alleles per locus ranged from 3 to 16 and the observed and expected heterozygosity ranged from 0.286 to 0.650 and 0.504 to 0.906, respectively. Five loci were found to be significantly deviated from the Hardy-Weinberg equilibrium and three pairs showed evidence of linkage disequilibrium. Most loci are highly informative for population genetics and linkage analyses according to their polymorphism information content (> 0.5) and will be useful for increasing our understanding of the wild population structure and developing a sustainable fishery management.

Keywords: Microsatellites, Next Generation Sequencing, *Panopea globosa*, Cortés Geoduck

Aislamiento y caracterización de nuevos marcadores microsatelitales en la almeja generosa (*Panopea globosa*)

RESUMEN. La almeja generosa *Panopea globosa* es una especie infáunica longeva y de gran tamaño que mantiene una pesquería creciente en la costa del Noroeste de México. A pesar de su demanda creciente en los mercados asiáticos, se conoce muy poco acerca de su biología. Con la finalidad de proveer nuevos marcadores genéticos para la caracterización de poblaciones silvestres, se desarrollaron nueve marcadores microsatelitales nuevos (con patrones repetidos de di-, tri-, y tetranucleotídicos) utilizando secuenciación genómica aleatoria con tecnología de secuenciación de siguiente generación (Illumina). El número de alelos por locus varió de 3 a 16 y los valores de heterocigosidad observada y esperada variaron de 0.286 a 0.650 y 0.504 a 0.906, respectivamente. Cinco microsatelites se desvían del equilibrio de Hardy-Weinberg y tres pares de microsatélites mostraron evidencia de desequilibrio de ligamiento. La mayoría de los loci son altamente informativos para estudios poblacionales y análisis de ligamiento de acuerdo con su contenido de información de polimorfismos (> 0.5) y serán útiles para incrementar el conocimiento de la estructura genética de las poblaciones silvestres de esta almeja y para coadyuvar en su pesquería sustentable.

Palabras clave: Microsatélites, Secuenciación de siguiente generación, *Panopea globosa*, Almeja generosa.

Bisbal-Pardo, C. I., M. Á. Del Río-Portilla, A. Y. Castillo-Paéz & A. Rocha-Olivares. 2016. Isolation and characterization of new microsatellite markers for the Cortés Geoduck (*Panopea globosa*). *CICIMAR Oceánides*, 31(1): 17-22.

INTRODUCTION

The Cortés Geoduck, *Panopea globosa* (Dall, 1898) is a marine bivalve endemic of the west coast of México. In the past, its distribution was thought to be limited to the Gulf of California but genetic analyses recently revealed its presence in Bahía Magdalena, in the west coast of Baja California Sur (Suárez-Moo *et al*., 2013). The commercial exploitation of Geoduck clams (*P. globosa* and *P. generosa*) in the northwest coast of México is mostly sustained by *P. globosa* and has significant commercial value, being mostly exported to Asian markets. This fishery started in 2002, however many aspects of Cortés Geoduck biology remain unknown.

Microsatellite markers are very useful in population genetic studies because they are highly polymorphic and abundant in the genome (Schlotterer, 2004). Originally, their isolation and characterization involved screening of genomic libraries obtained by cloning, which was expensive and labor-intensive. At present, shotgun genome sequencing, using next generation sequencing (NGS) technologies, has become the prevailing experimental approach and has helped to identify a large number of microsatellites in the several non-model species (Abdelkrim *et al*., 2009; Lance *et al*., 2013). In *Panopea globosa*, 24 tetranucleotide microsatellite loci were developed using the Roche 454 pyrosequencing platform (Cruz-Hernández *et al.*, 2014). Later, some of them were used in a genetic survey of the species, but a few had to be excluded because of problems with heterozygosity and linkage (Munguia-Vega *et al.,* 2015). In this study we aim to increase the number of informative microsatellites with varying repeat motifs (di-, tri- and tetranucleotide) using an alternative NGS platform (i.e., Illumina). These markers will add to the genetic resources available for the species and will be useful to advance the biological and ecological knowledge of this valuable fishing

Fecha de recepción: 05 de abril 2016

Fecha de aceptación: 29 de abril de 2016

resource, which is essential to design appropriate fishery management strategies and ensure its conservation.

MATERIALS AND METHODS

DNA extraction and whole genome sequencing

Genomic DNA was extracted from siphon tissues of three specimens, two of them collected in San Felipe (upper Gulf of California) and one in Bahía Magdalena (West coast of Baja California Sur), using the DNeasy blood and tissue kit (QIAGEN, Hilden, Germany). Extractions yielded over 50 ng.μl^{-1} of high quality (A260/A280 > 1.80) genomic DNA. Separate shotgun DNA libraries (~500 bp insert) were prepared with the TruSeq NanoTM kit (Illumina, San Diego, CA) and each sequenced in 1/12 lane of a HiSeq 2500 ultra-high-throughput sequencer (2x100 base-pair run, Illumina, San Diego, CA).

Isolation of microsatellite markers and primer design

Individual DNA sequences ("reads" henceforth) from the three libraries were subject to *de novo* assembly into contigs using CLC Genomics Workbench 7.0.3. Microsatellite search and PCR primer design were conducted using resulting contigs with the program Msatcommander (Faircloth, 2008). We searched for di-, tri- and tetranucleotide perfect microsatellite motifs with a minimum of eight, eight and six repeats, respectively. Additional search criteria included: primer length range 19-24 bp, melting temperature Tm 54-60ºC with matching Tm for each pair, regions with minimum 5X coverage and product sizes 140-400 bp. From the candidate microsatellite loci, we selected thirty (10 each size category) to test experimentally.

Amplification and genotyping

Experimental screening consisted in the optimization of PCR conditions in a small set of clams (n= 10). Markers successfully optimized were subsequently scored in a sample of 35 specimens (11 from San Felipe - 31º 01' 30'' N 114º 50' 27'' W - and 12 from Puerto Peñasco - 31º 19' 05'' N 113º 32' 12'' W -, in the Gulf of California and 12 from Bahía Magdalena - 24º 39' 10'' N, 112º 01' 47'' W -, in the western Pacific coast of Baja California Sur) to assess polymorphism using a fluorescently labeled universal M13 primer for amplification (Schuelke, 2000).

PCR reactions (15 µl) contained 1 X PCR Perkin Elmer Buffer (10mM Tris HCl, 50mM KCl, 2mM $MgCl_2$), 0.2 mM each dNTP, 1 U Taq DNA polymerase, 0.04 µM of the unlabeled M13-tailed forward primer, 0.6 µM of the fluorescently-labeled M13 primer, 0.04 µM of the reverse primer and 20ng of genomic DNA. According to de Arruda *et al*. (2010) thermal cycling was performed with an initial 5 min at 94ºC, followed by 40 cycles of 30 s at 94ºC, 30 s at the optimized melting temperature (Tm) (55.6 – 66 ºC) and 45 s at 72ºC. At this point the fluorescently-labeled M13 primer was added and additional 10 cycles of 30 s at 94ºC, 45 s at 53ºC and 45 s at 72ºC were performed, followed by final extension at 72ºC for 30 min. Genotyping was carried out using an ABI-3130xl automated DNA sequencer. Allele calling was conducted with the program Gene Marker 2.4.0 to score alleles and discrete alleles were assigned by FLEXIBIN (Amos *et al*., 2007). Finally, genotyping errors (stutters, allele dropout and typography) and the presence of null alleles were tested with MICRO-CHECKER v2.2.3 (Van Oosterhout *et al*., 2004).

Characterization of microsatellite loci

Estimates of the number of alleles per locus (Na), observed and expected heterozygosity (Ho and He) and polymorphism information content (PIC) were calculated with MStools (Park, 2001). We tested deviations from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) using ARLEQUIN v3.1 (Excoffier *et al*., 2005). Significance was adjusted for multiple statistical testing using Dunn-Sidák correction (Šidák, 1967).

RESULTS AND DISCUSSION

A total of 77,381,780 good quality reads were obtained from the sequencing platform. *De novo* assembly produced 736,851 contigs with an average length of 382 bp. The Illumina Hi-Seq 2500 is a high throughput platform with a large sequence yield per run and the paired-end option that allows for a higher success and reliability in *de novo* assembly of contigs (Bahassi & Stambrook, 2014)=<B. The same strategy has been successfully used in other non-model species (Nunziata *et al.*, 2012; O'Bryhim *et al.,* 2012). In total, 4,956 di- , 1,647 tri- and 1,983 tetranucleotide microsatellites were identified and primers could be designed for 18%, 12% and 23% of each them, respectively. Of the 30 loci tested, we successfully amplified and scored nine polymorphic loci resulting in a yield of 30%. In the first step, ten microsatellite loci were discarded due to non-reliable or non- target amplification. Amplification problems in microsatellite loci are common in bivalves (Selkoe & Toonen, 2006), this is likely to be a result of high frequency of repetitive elements in mollusks as revealed by comparative genomic analyses (McInerney *et al.*, 2011). A fraction of microsatellites was discarded because of inconsistent amplification and genotyping (i.e., un-interpretable chromatograms), possibly related to interference of the M13 tail during primer annealing, as found elsewhere (Guichoux *et al.*, 2011).

We developed nine novel polymorphic microsatellites with varying repeat motifs (one di-, three tri- and five tetranucleotide, Table 1). At present, only tetranucleotide loci have been described for *Pano-*

Table 1. Polymerase chain reaction primers and levels of polymorphism of novel microsatellite loci developed for *Panopea globosa* using next generation sequencing.

Locus	Primer sequence (5'-3')	Tm	Motif	n	Na	Allelic range	HO	HE	PHWE	PIC	GenBank accession
Pglo2_6	F:GCCCACATCAACTTGAACCC R:ACGAAACAACCAACATCGTCA	66	(AT)8	24	3	280-288	0.583	0.504	0.390	0.391	KX355764
Pglo3_2	F:ACGGAATTAGGAATATGCGCAC R:ACCGAATATCACACACCATGC	63	(AAC)10	29	16	239-341	0.517	0.834	0.000*	0.808	KX355765
Pglo3_4	F: GGTTTCATCGGCCACTGAATG R:AACTCAAACACTGTCCTGCC	62	(AAT)16	21	5	236-248	0.429	0.587	0.010	0.538	KX355766
Pglo3_5	F:TTCCCTGTATCGCGTTATGG R: ACATTGCAGTTTGATTGGGC	62	(AAC)9	21	8	336-378	0.286	0.827	0.000*	0.781	KX355767
Pglo4_1	F: CTCTCGCCGGAATTGGTATG R:TCACAACCCCACTAAACCGC	55.6	(AAAG)8	20	7	260-328	0.650	0.700	0.600	0.629	KX355768
Pglo4_2	F:AACAGCGTTAATTGAGGGCG R:ATGATCGTGAAATCGCCAGC	68	(ACGC)8	28	14	317-445	0.607	0.906	0.000*	0.881	KX355769
Pglo4_3	F:TGCCCATTGCTATCACGCAT R: AAGCACGATCATTACTCCGG	55.6	(AAAT)7	28	6	261-285	0.464	0.747	0.003*	0.688	KX355770
Pglo4_5	F:GTCTGTCAGCCCTCAATAGC R:TCACTGCTCATGCATCCTTCA	66	(AATC)7	15	7	247-275	0.533	0.809	0.029	0.750	KX355771
Pglo4_6	F:GTGAAAATTGCACAACGATAAATGT R:GCCACCACAATGGTAAGCTG	66	(ACGC)8	23	10	349-385	0.565	0.886	0.000*	0.852	KX355772

Abbreviations: Tm (°C): annealing temperature, n: sample size; Na: number of alleles; H_O: observed heterozygosity; H_E: expected heterozygosity; P_{HWE}: Hardy-Weinberg equilibrium test p-value (*, $p < 0.005$); PIC: polymorphism information content.

pea globosa. The availability of loci with smaller repeat motifs and higher mutation rates (Chakraborty *et al.*, 1997; Estoup & Cornuet, 1999) may help to uncover additional genetic variability for increased power of genetic analyses. The number of alleles ranged from 3 to 16 (mean number of alleles per locus = 8.56), and heterozygosity estimates between 0.286 and 0.650 (mean = 0.515), whereas the PIC varied between 0.391 and 0.881 (mean = 0.702). Five loci (Pglo3_2, Pglo3_5, Pglo4_2, Pglo4_3 and

Pglo4_6) had significant heterozygote deficiencies (after Dunn-Sidák correction, $p < 0.05$) and were flagged by MICRO-CHECKER for the presence of null alleles. A large number of loci deviating from HWE was also found in the tetranucleotide loci previously developed for *P. globosa* (Cruz-Hernández *et al.*, 2014). The results of MICRO-CHECKER are also in agreement with the high incidence of null alleles in mollusk microsatellites (Hedgecock *et al.*, 2004; Becquet *et al*., 2009; McInerney *et al*., 2011). In addition, three pairs of loci (Pglo2_6-Pglo4_2, Pglo3_2-Pglo4_ 6, Pglo4_5-Pglo4_6) were found to be in linkage disequilibrium ($p < 0.001$). A possible Wahlund effect in the HW and linkage disequilibria cannot be rejected since clams from the Gulf of California have been shown to be genetically differentiated from those in the Pacific (Munguia-Vega *et al.*, 2015). The additional microsatellite loci described herein will prove useful in population genetics studies of *P. globosa* involved in the sustainable management of wild populations or in aquaculture endeavors.

ACKNOWLEDGMENTS

This research was funded by COFUPRO grant RGAC-CES-2013-031 to ARO. The first author benefited from a graduate fellowship from CONACYT to support her M. Sc. program in Marine Ecology at Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE).

REFERENCES

Abdelkrim, J., B. Robertson, J. A. Stanton & N. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, 46 (3): 185-192. https://doi.org/10.2144/000113084

Amos, W., J. I. Hoffman, A. Frodsham, L. Zhang, S. Best & A. V. S. Hill. 2007. Automated binning of microsatellite alleles: problems and solutions. *Mol. Ecol. Notes*, 7 (1): 10-14. https://doi.org/10.1111/j.1471-8286.2006.01560.x

Bahassi, E. & P. J. Stambrook. 2014. Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis,* 29 (5): 303-310. https://doi.org/10.1093/mutage/geu031

Becquet, V., I. Lanneluc, B. Simon-Bouhet & P. Garcia. 2009. Microsatellite markers for the Baltic clam, *Macoma balthica* (Linne, 1758), a key species concerned by changing southern limit, in exploited littoral ecosystems. *Conserv. Genet. Resour.,* 1 (1): 265-267. https://doi.org/10.1007/s12686-009-9065-0

Chakraborty, R., M. Kimmel, D. N. Stivers, L. J. Davison & R. Deka. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *P. Natl. Acad. Sci. USA*, 94 (3): 1041-1046. https://doi.org/10.1073/pnas.94.3.1041

Cruz-Hernández, P., A. Munguía-Vega, I. Leyva-Valencia, F. Lucero-Burquez & D. B. Lluch-Cota. 2014. Development of 24 tetra-nucleotide microsatellite markers in Cortes Geoduck *Panopea globosa* by next-generation sequencing. *Conserv. Genet. Resour.*, 6 (3): 531-533. https://doi.org/10.1007/s12686-014-0172-1

de Arruda, M. P., E. C. Gonçalves, M. P. C. Schneider, A. L. D. C. da Silva & E. Morielle-Versute. 2010. An alternative genotyping method using dye-labeled universal primer to reduce unspecific amplifications. *Mol. Biol. Rep.*, 37 (4): 2031-2036. https://doi.org/10.1007/s11033-009-9655-7

Estoup, A. & J. M. Cornuet. 1999. Microsatellite evolution: inferences from population data. 49-65, *In*: Goldstein, D. B. , Schlötterer, C. (eds.). *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford.

Excoffier, L., G. Laval & S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolut. Bioinform. Online*, 1: 47-50. https://doi.org/10.1177/117693430500100003

Faircloth, B. C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.*, 8 (1): 92-94. https://doi.org/10.1111/j.1471-8286.2007.01884.x

Guichoux, E., L. Lagache, S. Wagner, P. Chaumeil, P. Leger, O. Lepais, C. Lepoittevin, T. Malausa, E. Revardel, F. Salin & R. J. Petit. 2011. Current trends in microsatellite genotyping. *Mol. Ecol. Resour.*, 11 (4): 591-611. https://doi.org/10.1111/j.1755-0998.2011.03014.x

Hedgecock, D., G. Li, S. Hubert, K. Bucklin & V. Ribes. 2004. Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *J. Shellfish Res.*, 23 (2): 379-385.

Lance, S. L., C. N. Love, S. O. Nunziata, J. R. O'Bryhim, D. E. Scott, R. W. Flynn & K. L. Jones. 2013. 32 species validation of a new Illumina paired-end approach for the development of microsatellites. *PLoS One*, 8 (11): e81853. https://doi.org/10.1371/journal.pone.0081853

McInerney, C. E., A. L. Allcock, M. P. Johnson, D. A. Bailie & P. A. Prodohl. 2011. Comparative genomic analysis reveals species-dependent complexities that explain difficulties with microsatellite marker development in molluscs. *Heredity,* 106 (1): 78-87. https://doi.org/10.1038/hdy.2010.36

Munguia-Vega, A., I. Leyva-Valencia, D. B. Lluch-Cota & P. Cruz-Hernández. 2015. Genetic structure of the Cortes Geoduck *Panopea globosa* Dall, 1898, from the Mexican Northwest. *J. Shellfish Res.*, 34 (1): 153-161. https://doi.org/10.2983/035.034.0119

Nunziata, S. O., J. D. Karron, R. J. Mitchell, S. L. Lance, K. L. Jones & D. W. Trapnell. 2012. Characterization of 42 polymorphic microsatellite loci in *Mimulus ringens* (Phrymaceae) using Illumina sequencing. *Am. J. Bot.*, 99 (12): E477-E480. https://doi.org/10.3732/ajb.1200180

O'Bryhim, J., J. P. Chong, S. L. Lance, K. L. Jones & K. J. Roe. 2012. Development and characterization of sixteen microsatellite markers for the federally endangered species: *Leptodea leptodon* (Bivalvia: Unionidae) using paired-end Illumina shotgun sequencing. *Conserv. Genet. Resour.,* 4 (3): 787-789. https://doi.org/10.1007/s12686-012-9644-3

Park, S. 2001. *MStools v 3 (Excel spreadsheet toolkit for data conversion). Smurfit Institute of Genetics. Trinity College, Dublin.*

Schlotterer, C. 2004. The evolution of molecular markers - just a matter of fashion? *Nat. Rev. Genet.*, 5 (1): 63-69. https://doi.org/10.1038/nrg1249

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.*, 18 (2): 233-234. https://doi.org/10.1038/72708

Selkoe, K. A. & R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.*, 9 (5): 615-629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Šidák, Z. K. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc.,* 62 (318): 626–633. https://doi.org/10.1080/01621459.1967.10482935

Suárez-Moo, P. J., L. E. Calderon-Aguilera, H. Reyes Bonilla, G. Díaz-Erales, V. Castañeda-Fernandez-de-Lara, E. A. Aragón-Noriega & A. Rocha-Olivares. 2013. Integrating genetic, phenotypic and ecological analyses to assess the variation and clarify the distribution of the Cortes Geoduck (*Panopea globosa*). *J. Mar. Biol. Assoc. U. K.*, 93 (3): 809–816. https://doi.org/10.1017/S0025315412001464

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills & P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes,* 4 (3): 535-538. https://doi.org/10.1111/j.1471-8286.2004.00684.x