López-Márquez, V.; Alfaya, J. E. F.; García-Jiménez, R.; Bigatti, G. y Machordom, A. 2013. Isolation and characterisation of microsatellite loci for the southern geoduck, /Panopea abbreviata/ (Valenciennes, 1839) through 454 pyrosequencing.

Included in: Ahanchédé, A.; Alfaya, J. E. F.; Andersen, L. W.; Azam, D.; Bautista, M. A. M.; Besnard, A. L.; Bigatti, G.; Bouetard, A.; Coutellec, M. A.; Ewedje, E.; Fuseya, R.; Garcia-Jimenez, R.; Haratian, M.; Hardy, O. J.; Holm, L. E.; Hoy, C. W.; Koshimizu, E.; Loeschcke, V.; Lopez-Marquez, V.; Machado, C. A.; Machordom, A.; Marchi, C.; Michel, A. P.; Micheneau, C.; Mittapalli, O.; Nagai, T.; Okamoto, N.; Pan, Y.; Panitz, F.; Safaie, N.; Sakamoto, T.; Sharifnabi, B.; Tian, E. W. y Yu, H. 2013. Permanent Genetic Resources added to Molecular Ecology Resources Database 1 August 2012-30 September 2012. Molecular Ecology Resources 13(1): 158-159 (2012).

Isolation and characterisation of microsatellite loci for the southern geoduck *Panopea abbreviata* (Valenciennes, 1839) through 454 pyrosequencing

Violeta López-Márquez[1a], José E. F. Alfaya[2a], Ricardo García-Jiménez[1], Gregorio Bigatti[2] and Annie Machordom[1].

[1]Museo Nacional de Ciencias Naturales (MNCN-CSIC). José Gutiérrez Abascal, 2. 28006 Madrid, Spain.

[2]LARBIM, Centro Nacional Patagónico (CENPAT-CONICET). Bvd. Brown 2915, U9120ACV Puerto Madryn, Chubut, Argentina.

[a]These authors contributed equally to this work.

Keywords: *Patagonian geoduck, artisanal fisheries, 454 GS-FLX Titanium pyrosequencing, population structure*

Correspondence: Annie Machordom. E-mail: annie@mncn.csic.es. Telephone: +34 914111328. Fax number: +34 915645078.

Short title: *Panopea abbreviata* microsatellite markers

Abstract

We have isolated the first, polymorphic microsatellite loci (21 in total) for the geoduck clam *Panopea abbreviata* (Valenciennes, 1839) from San Matías gulf (Patagonia, Argentina), using 454 GS-FLX Titanium pyrosequencing. We also developed conditions for amplifying these markers in 5 multiplex and 7 individual reactions. Four to 23 alleles were detected per locus across the 25 samples analysed. Observed and expected heterozygosities ranged from 0.235 to 0.985 and from 0.528 to 0.937, respectively. In the sampled population, only one locus deviated from Hardy–Weinberg equilibrium. These markers are useful resources for future population structure studies in this artisanal fishing species.

The genus *Panopea* Ménard, 1807 (Bivalvia, Hiatellidae), has nine species, and its populations naturally occur worldwide (Straus *et al.* 2008). These bivalves, called geoducks, have one of the longest lifespans among exploited animals (Morsan *et al.* 2010). There is lucrative commerce based on geoduck fisheries, mostly from the North Pacific Ocean. Probably due to this fact, the Pacific geoduck *Panopea generosa,* previously known as *P. abrupta* (Vadopalas *et al.* 2010), was the first species of the genus to have its genetic structure analysed (Vadopalas and Bentzsen 2000; Vadopalas *et al*. 2004).

Here, we focused on the southern Patagonian geoduck *Panopea abbreviata* (Valenciennes, 1839), the largest bivalve found along the south-occidental Atlantic coasts, with a distribution range from Rio de Janeiro, Brazil (23ºS) to Nuevo gulf, Argentina (43ºS) (Scaravino 1977, Alfaya *et al.* in press). These geoducks live from shallow waters to depths of 75 m, and are buried (up to 70 cm) in sand and muddy sediments (Ciocco 2000). It has been an incipient resource for the artisanal fisheries in the Patagonian gulfs since 1999 (Ciocco 2000; Ciocco *et al*. 2001). However, there is not much basic ecological or biological information about *P. abbreviata*, and there is no regulation for its management. Recently, its reproductive cycle was described as having a continuous gametogenetic cycle (Van der Molen *et al*. 2007; Zaidman *et al*. 2012). This lack of resting in reproduction coupled with a pelagic larval stage could indicate a lack of differentiation, but some differences in growth and age classes were reported among populations (Morsan *et al*. 2010).

To disentangle the basis of such differences, including factors that may influence dispersal and gene flow (Acevedo *et al.* 2009), and to provide a new tool that will aid in the sustainable management of this species, we have developed the first series of

microsatellite markers for *P. abbreviata*. Thus, we isolated and screened the microsatellites using high-throughput sequencing techniques. Multiplex-enriched libraries and 454 GS-FLX Titanium pyrosequencing are powerful tools for the isolation of new markers in unknown genomes (Martin *et al.* 2010). This procedure has been readily and successfully applied to a large variety of taxonomic groups (Malausa *et al.* 2011).

Genomic libraries were constructed at the Cornell Evolutionary Genetics Core Facility (EGCF, USA). Total DNA was extracted from the siphon tissue of one *P. abbreviata* specimen from San Matías gulf (Patagonia, Argentina), using the BioSprint 15 DNA Blood Kit (Qiagen, Hilden, Germany), according to manufacturer's tissue protocol.

Five micrograms of extracted genomic DNA were completely digested with a restriction enzyme (five-base cutter) that generated blunt-end fragments. Linkers were ligated to the digested DNA, and the resulting fragments were enriched for microsatellites by hybridization to and magnetic capture of biotinylated repeat probes (representing two unique dimers –GT and TC; five unique trimers –TTC, GTA, GTG, TCC and GTT; and five unique tetramers –TTTC, GATA, TTAC, GATG and TTTG). Enriched genomic fragments captured by streptavidin-coated magnetic beads were then amplified by PCR, ligated to Roche/454 Titanium Multiplex Identifier (MID) adapters and size fractionated in an agarose gel.

Libraries with unique adapters were pooled, and sequences were generated with Roche/454 GS FLX Titanium reagents, protocols and hardware. MID-sorted 454 reads were trimmed of adapter sequences and assembled with SeqMan Pro (DNASTAR).

Consensus files and singleton reads were exported as FASTA files, and simple repeats were detected with MsatCommander software (Rozen and Skaletsky 2000; Faircloth 2008), with a limit of at least 8 perfect motif repeats.

We primarily focused on finding tetranucleotide repeats as they typically provide clear allele assignment. From the 1004 perfect tetranucleotide repeats obtained, 208 showed the possibility of designing primers under the requested conditions (e.g., length 20±2 bases, melting temperature 60±2 ºC, GC content between 30% and 70%). Then, we discarded the ligated microsatellites (that had more than one group of repeats per contig), leaving 140 viable microsatellites, with 21 different motifs. To have a good representation of all motifs, when possible, we selected at least 2 of each. Finally, we tested 41 potential microsatellites markers in 7 samples from 5 localities from the Patagonian gulfs using a nested PCR protocol modified from Schuelke (2000). Twenty-two microsatellites markers produced clear electropherogram patterns in these 7 samples and were selected for multiplex PCR and genotyping optimization.

Multiplex PCRs were performed in a total volume of 10 µl, which included approximately 1 ng of DNA, 1X Qiagen Multiplex PCR Master Mix (Qiagen, Hilden, Germany), and $MgCl_2$ at a final concentration of 3 mM. Primer concentrations ranged from 0.2 to 0.8 µM (0.6 for Pa19 locus primers, 0.8 for Pa45 locus primers and 0.2 for the all other primers). The forward primer from each primer pair was fluorescently 5' end labelled with 6-FAM, NED, VIC or PET, while the reverse primers were pig-tailed with 5′-GTTTCTT-3′ (Brownstein, 1996) to facilitate genotyping. The cycling profile began with an enzyme activation step at 95 °C for 15 min (Qiagen Multiplex PCR Kit specifications), followed by 35 cycles at 94 ºC for 30 s, 56 ºC for 90 s (except for three loci: Pa17, Pa31 and Pa33, for which the annealing temperature was 52 °C), 72 ºC for

60 s and a final extension at 60 ºC for 30 min. A Mastercycler gradient thermocycler (Eppendorf AG, Hamburg, Germany) was used for all reactions.

Fluorescently labelled PCR products were run on an ABI PRISM 3730 DNA Sequencer (Applied Biosystems) with the GeneScan-500 (LIZ) internal size standard and analysed with the GeneMapper software (Applied Biosystems).

From the 22 loci surveyed, 21 loci were successfully amplified and genotyped in 25 specimens from San Matías gulf (40º50'S, 65º04'W; Patagonia, Argentina). Overall, 1 tetraplex, 2 triplex, 2 diplex and 7 monoplex were optimized to reduce the number of PCRs performed to 12 (Table 1). Two of the amplified and genotyped microsatellites were pentanucleotide repeats, while 19 were tetranucleotide repeats. These microsatellite markers were then checked for accordance to Hardy-Weinberg equilibrium (HWE) and the presence of linkage disequilibrium, using GENEPOP 4.1 (Rousset 2008) with a Markov chain method (5000 dememorization steps, 100 batches and 5000 iterations per batch). Only one among the 210 tests was significant for linkage disequilibrium, but no significance was found after sequential Bonferroni correction for multiple testing (Holm 1979; Rice 1989). One locus (Pa 25) showed a significant departure from HWE ($p = 0.0016 \pm 0.0003$), due to heterozygoty deficit, which could be because of the presence of null alleles, as was found using MICRO-CHECKER (Van Oosterhout *et al*. 2004). This program found that all loci lacked apparent errors of scoring, large allele dropout and the existence of null alleles, except for the aforementioned Pa25. Null alleles were also identified at locus Pa17 by MICRO-CHECKER, but not at significant frequencies. Basic parameters of genetic variability were calculated using GenAlex (Peakal and Smouse 2006). Allele richness ranged from 4 to 23, while observed and expected heterozygosities ranged from 0.235 to 0.985 and

131 0.528 to 0.937, respectively (Table 1). The lowest observed heterozygosity value was
132 for the locus that showed the heterozygosity deficit (P25). Excluding this locus, the
133 heterozygosity indices obtained here for *P. abbreviata* were higher than those observed
134 in the 30 specimens of *P. generosa* (called before *P. abrupta*) analysed by Vadopalas
135 and Bentzen (2000), even though the 5 microsatellite markers they isolated had a higher
136 number of alleles per locus (7 to 33).

137 These markers will be valuable tools for population genetic studies that examine the
138 dynamics, connectivity and variability of populations of *Panopea abbreviata*, thus
139 helping managers to conserve and handle this artisanal fishery resource.

140

141 **Acknowledgments**

142 We would like to thank Steve Bogdanowicz from the Cornell EGCF for his technical
143 assistance during microsatellite isolation. Melinda Modrell revised the English. This
144 research was funded by the Spanish Ministry of Foreign Affairs through the AECID
145 grants A/023484/09 and A/032441/10.

146

147 **References**

148 Alfaya JEF, Bigatii G, Machordom A (in press). Mitochondrial and nuclear markers
149 reveal a lack of genetic structure in the entocommensal nemertean *Malacobdella*
150 *arrokeana* in the Patagonian gulfs. *Helgoland Marine Research*. DOI:
151 10.1007/s10152-012-0326-z

152 Acevedo I, Bloor P, Toledo C, Calvo M, Machordom A (2009) Development of
153 tetranucleotide microsatellite markers for the cushion star, *Asterina gibbosa,* and
154 cross-species amplification. *Molecular Ecology Resources,* **9**, 274-277.

Ciocco NF (2000) Almeja panopea, un nuevo recurso pesquero para el Mar Argentino. *Infopesca Internacional*, **6**, 36-39.

Ciocco NF, de Garin N, Diaz MA, Vera R, Mazzanti R, Monsalve MA, Herrera G, Sollazo S, Serda A, Diaz D, Signorelli C, Lopez J, Ascorti J, Diaz R, Bazterrica MC, Escati G, Real L (2001). Relevamiento de bancos de moluscos bivalvos de interés marisquero en el golfo San José. Resultados de la campaña Sanjo/01. Acta Complementaria Convenio Provincia del Chubut- CENPAT-CONICET Ley 3315. Inf. nº 11 del LAPEMAR, Laboratorio de Peces y Mariscos de Interés Comercial (CENPAT): 1-69.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, **8**, 92-94.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics,* **6**, 65–70.

Malausa T, Gilles A, Meglécz E, Blanquart H, Duthoy S, Costedoat C, Dubut V, Pech N, Castagnone-Sereno P, Délye C, Feau N, Frey P, Gauthier P, Guillemaud T, Hazard L, Le Corre V, Lung-Escarmant B, Malé P-J, Ferreira S. Martin, J-F (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources*, **11**, 638-644.

Martin J-F, Pech N, Meglécz E, Ferreira S, Costedoat C, Dubut V, Malausa T, Gilles A (2010) Representativeness of microsatellite distributions in genomes, as revealed by 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*, **11**, 560.

178 Morsan E, Zaidman P, Ocampo-Reinaldo M, Ciooco N (2010) Population structure,
179 distribution and harvesting of southern geoduck, *Panopea abbreviata*, in San
180 Matías Gulf (Patagonia, Argentina). *Scientia Marina*, **74**, 764-772.
181 Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population
182 genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288-295.
183 Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.
184 Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software
185 for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
186 Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist
187 programmers. In: Bioinformatics Methods and Protocols: Methods in Molecular
188 Biology (eds Krawetz S, Misener S). Humana Press, Totowa, NJ.
189 Scarabino V (1977) Moluscos del Golfo San Matías (Prov. Río Negro, Argentina).
190 Inventario y claves para su identificación. *Comunicaciones de la Sociedad*
191 *Malacológica de Uruguay*, **IV**, 177-297.
192 Schuelke M (2000) An economic method for the fluorescent labelling of PCR
193 fragments. *Nature Biotechnology,* **18**, 233–234.
194 Straus KM, Crosson LM, Vadopalas B (2008) Effects of geoduck aquaculture on the
195 environment: A synthesis of current knowledge. Washington Sea Grant Technical
196 Report WSG-TR 08-01.
197 Vadopalas B, Bentzen P (2000) Isolation and characterization of di- and tetranucleotide
198 microsatellite loci in geoduck clams, *Panopea abrupta*. *Molecular Ecology*, **9**,
199 1435-1436.

Vadopalas B, LeClair LL, Bentzen P (2004) Microsatellite and allozyme analyses reveal few genetic differences among spatially distinct aggregations of geoduck clams (*Panopea abrupta*, Conrad 1849). *Journal of Shellfish Research,* **23**, 693–706.

Vadopalas B, Pietsch TW, Friedman CS (2010) The proper name for the geoduck: Resurrection of *Panopea generosa* Gould, 1850, from the synonymy of *Panopea abrupta* (Conrad, 1849) (Bivalvia: Myoida: Hiatellidae). *Malacologia*, **52**, 169-173.

Van der Molen S, Kroeck M, Ciocco N (2007) Reproductive cycle of the southern geoduck clam, *Panopea abbreviata* (Bivalvia: Hiatellidae), in north Patagonia, Argentina. *Invertebrate Reproduction & Development*, **50**, 75-84.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Zaidman PC, Kroeck MA, Kissner EMO, Morsan EM (2012) Reproductive pattern of Southern geoduck, *Panopea abbreviata*, at El Sotano (San Matias Gulf, Patagonia, Argentina). *Marine Biology Research*, **8**, 172 -181.

Table 1 Characterisation of 21 microsatellite loci in *Panopea abbreviata.* Primer sequences, repeat motifs, clone size (bp), alleles size range (bp). PCR reaction indicates multiplex (loci with the same number in this column were amplified and genotyped together) or individual reactions. Na= number of alleles, H_O= observed heterozygosity, H_E= expected heterozygosity. GenBank accession numbers for each microsatellite. *indicates significant deviation from Hardy-Weinberg equilibrium.

Locus name	Primer sequences (5'-3')	Repeat motif	Clone size	PCR reaction	Alleles range	Na	H_O	H_E	Accession number
Pa6	F VIC-CATGTTTACAGAAGTTAGGC R ACAGCAAGATGTTGAAACT	$(TACTC)_{11}$ $(TACCC)_5$ $(TACTC)_2$	329	1	318-368	10	0.840	0.866	JX416866
Pa8	F 6-FAM-ATAACATGTAAATGTATCATTAGAG R TATTTGACGTTAGGACGTTT	$(CCATT)_8$	163	2	154-179	6	0.680	0.747	JX416867
Pa13	F 6-FAM-CGTTTACTCAAACATGGTAT R TGAACATCTTTCTATAATTTTATCT	$(CAGA)_{12}$	143	3	123-179	11	0.760	0.859	JX416868
Pa14	F 6-FAM-AAAGGCAAGGTGGCTTGT R TTTCACGGATAGTGAATTTCG	$(CATA)_{11}$	166	4	161-315	22	0.880	0.924	JX416869
Pa16	F 6-FAM-CAATAGCTCGCCTTATTAC R CTGACCGTCTGATAGCTC	$(GTTT)_8$	131	5	130-154	5	0.440	0.607	JX416870
Pa17	F PET-TTTGTAATATGACGTTCTTG R AATAAAACGTTCACAGAGAC	$(TTAC)_9$ $(TTAA)_4$ TTAC TTAA $(TTAC)_3$	246	6	203-679	14	0.720	0.899	JX416871
Pa18	F VIC-CGTTTGTTCTAGTGTTGAG R GTACACCTGTAAATCAGACC	$(TCCA)_{12}$	367	7	356-420	15	0.840	0.910	JX416872
Pa19	F 6-FAM-ATTTATAACCTCCATAATGC R ACAAACACAATTTAATAACG	$(CATT)_{13}$	179	7	155-199	11	0.667	0.744	JX416873
Pa20	F PET-TTGGACTGAGTTATTAAAGG R CCATGAGACATGACATTG	$(CCGT)_8$ CCAT $(CCGT)_3$	252	1	241-261	4	0.560	0.573	JX416874
Pa23	F VIC-GACGTAATAATAGCGTGTTC R ATAAGACATTGAACGGAAG	$(CGCA)_{12}$	363	4	345-417	16	0.920	0.903	JX416875
Pa25	F PET-TTCTGTGTAAATGCATAGG	$(GCGT)_3$ GTGT GGTT $(GCGT)_8$	212	7	197-217	4	0.235	0.528*	JX416876

	R AGTAACGCGCTTATAGGT								
Pa27	F6-FAM-TTTCTGAGCTTGTATCTGTT R GTTATACGGAATAATCGTGA	$(CGTG)_{12}$	177	8	162-194	9	0.840	0.842	JX416877
Pa28	F PET-CTCGATGACATAATACGG R ACGTACTTTGTTTATAGGTCA	$(CAAT)_{10}$	205	9	189-217	7	0.680	0.693	JX416878
Pa29	F 6-FAM-GCCAGTATTGACTATTTTGT R GACGTGAACAATTAAGTAGAG	$(TGCG)_{9}$	194	1	194-226	7	0.520	0.576	JX416879
Pa31	F PET-TCCTTTATCCCTGTATTTG R TATTATTGTACTGTCATGCAC	$(GAGT)_{9}$	237	10	247-391	23	0.880	0.937	JX416880
Pa32	F NED-GACGTGACATAAAACAC R AATGTCACTTTTATTACTTC	$(CATT)_{9}$	115	9	119-179	11	0.880	0.874	JX416881
Pa33	F NED-TTAAATGCTGCATATTTTG R AATTTAAAATAGGCAATTACTC	$(TCTG)_{13}$	145	11	133-185	14	0.985	0.898	JX416882
Pa35	F 6-FAM-TTAGGAGATTGTAACAGAGC R ATTATACTACGCAGGAGGA	$(GGAT)_{9}$	161	9	157-189	8	0.640	0.774	JX416883
Pa36	F 6-FAM-TTCACCATCATCTTTAAAAC R GAAGAAGGACATTACATTGA	$(CGCA)_{12}$	161	12	136-164	8	0.800	0.812	JX416884
Pa39	F NED-AGTGACGTTACATTTACAGG R GTTCGATCATTTTAACATCT	$(GACT)_{3}$ $(GGCT)_{8}$	137	8	146-210	14	0.880	0.890	JX416885
Pa45	F VIC-GATTATTTAATAGTCTTAAATGG R CTAGTTAAAAGCAATGCTAA	$(AAAC)_{11}$	268	9	240-292	8	0.640	0.589	JX416886

220