Isolation and characterization of new microsatellite markers for the Pacific geoduck (Panopea generosa) using next generation sequencing

Aislamiento y caracterización de nuevos marcadores microsatélites para la almeja generosa (Panopea generosa) por medio de secuenciación de nueva generación

Celia Isabel Bisbal-Pardo¹, Miguel Ángel Del Río-Portilla², Ana Castillo-Páez¹ and Axayácatl Rocha-Olivares¹

¹Molecular Ecology Laboratory, Biological Oceanography Department, CICESE, Carretera Ensenada-Tijuana 3918, Zona Playitas, Ensenada, Baja California, 22860, México
²Genetics Laboratory, Aquaculture Department, CICESE, Carretera Ensenada-Tijuana 3918, Zona Playitas, Ensenada, Baja California, 22860, México
e-mail: arocha@cicese.mx


The Pacific geoduck, Panopea generosa Gould, 1850, is a large and long-lived infaunal clam, inhabiting sediments from the low intertidal to subtidal waters from Alaska, USA, to Baja California, México, and is the most important commercial geoduck (Bivalvia: Hiatellidae) in the northeast Pacific (Aragón-Noriega et al., 2012). Its commercial harvest started in 1970’s in the U.S. and Canada, and in early 2000 in Mexico (Aragón-Noriega et al., 2012). Commercial aquaculture operations started in the late 1990’s in the U.S. The species is sensible to over-exploitation because of its extended longevity, low recruitment, late sexual maturity, and slow growth (Bureau et al., 2002; Calderon-Aguilera et al., 2010; Orensanz et al., 2004; Sloan & Robinson, 1984). Therefore, assessing the structure and dynamics of wild populations is a prerequisite to address their potential for sustainable exploitation and to manage risk reduction in aquaculture (Straus et al., 2008). Ge-
nomic analyses using polymorphic loci are a powerful tool to investigate population genetic structure, and to assess levels of genetic variability, effective population size and extinction risk (Evans & Sheldon, 2008). Among molecular genetic markers, microsatellites show several advantages (Jarne & Lagoda, 1996) however, their isolation in a new species requires significant effort (time- and money-wise) and expertise (Zane et al., 2002). In addition, the genomic abundance of microsatellites is variable among taxa; and in some genomes, such as those of bivalves, they appear in low frequency (Peñarrubia et al., 2015). On the other hand, next generation DNA sequencing (NGS) technologies provide a cost-time-efficient means to isolate large number of microsatellites (Abdelkrim et al., 2009; Csencsics et al., 2010; Inoue et al., 2013; Lance et al., 2013).

**Panopea generosa**, formerly incorrectly named *Panopea abrupta* Conrad, 1849 (Vadopalas et al., 2010), has been subject to the development of seventeen species-specific microsatellites using traditional cloning methods (Kaukinen et al., 2004; Vadopalas & Bentzen, 2000; Vadopalas et al., 2004). Some of which have been used repeatedly in population genetics studies with variable and sometimes limited success (Miller et al., 2006; Suárez-Moo et al., 2016; Vadopalas et al., 2004; Vadopalas et al., 2012). Suárez-Moo et al. (2016) and Vadoplas et al. (2012) found genetic homogeneity between samples of Baja California and Washington and among cohorts of Washington, respectively. On the other hand, Miller et al. (2006) and Vadopalas et al. (2004) revealed genetic heterogeneity in populations of Canada and USA. In this study, we aim at producing additional genetic markers and complement the molecular toolbox of the Pacific geoduck to increase the power of future genetic studies along its distribution. This information will prove valuable for managing the exploitation of wild populations and help direct aquaculture efforts along its distribution area.

DNA of gill tissue was extracted from two fresh organisms, collected near Ensenada, Baja California, using the DNeasy blood and tissue kit (QIAGEN, Hilden, Germany), obtaining in excess of 50 ng/μl of high quality (A_{260}/A_{280} > 1.80) genomic DNA. All processes of library construction and Illumina sequencing were done as a described in Bisbal-Pardo et al. (2016). Bioinformatic analyses (quality control, ends trim, de novo assembly, microsatellite identification and primer design) were also carried out as described in Bisbal-Pardo (2014) and Bisbal-Pardo et al. (2016). For marker assessment, thirty microsatellite loci (9 di-, 10 tri- and 11 tetranucleotide) were selected (primer lengths ranging 19-24 bp, matching annealing temperature (Tm) between 54-60 °C, a minimum of 5X coverage and a product size of 140-400 bp, and their amplification was tested using the same PCR conditions of Bisbal-Pardo et al. (2016). Genotyping was performed using ABI-3130xl automated DNA sequencer. Alleles were scored with the program Gene Marker 2.4.0 and allele sizes were assigned to bins using FLEXBIN (Amos et al., 2007). We identified genotyping errors (stutters, allele dropout, typographical) and evaluated the presence of null alleles with MICRO-CHECKER (Van Oosterhout et al., 2004). Loci were scored in a set of 35 organisms, 7 from Hood canal (47°40’58.92” N 122°44’51.66” W) and 7 from Alden Bank (48°49’43.22” N 122°49’50.6” W), Washington, EUA; 7 from Coronado Islands (32°25’00”N 117°15’00” W), 7 from San Quintín (30°23’22.02” N 115°54’47.2” W) and 7 from Santa Rosalíta (28°40’00” N 114°15’57” W), Baja California, Mexico. We estimated the number of alleles per locus (k), observed and expected heterozygosities (Ho and He) and polymorphic information content (PIC) using Mstools (Park, 2001). Next, we calculated the deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) with ARLEQUIN 3.1 (Excoffier et al., 2005). Significance was adjusted for multiple testing using Dunn-Šidák correction (Šidák, 1967). We obtained a total of 77,475,634 reads from NGS, after the trimming, the 0.76% was discarded and 7.52% were corrected resulting in reads of 99.5 bp of the average length. We identified 8,060 di-, 3,146 tri- and 2,830 tetranucleotide microsatellites in a total of 868,521 contigs of 443 bp average length, N50 was 461 bp and average coverage was 11.86 reads. Less than 26% of markers were suitable for PCR primer design. The great disparity in the number of microsatellite loci identified bioinformatically and those amenable to primer design has been reported repeatedly (Castoe et al., 2012; Castoe et al., 2010; Csencsics et al., 2010).

Of the 30 loci tested, only eight were consistently and accurately genotyped (Table 1). The yield obtained in this study (27%), defined as the fraction of microsatellite loci successfully genotyped from the total experimentally tested, is similar to others obtained from mollusk using NGS (43%, An & Lee, 2012; 35%, Cruz-Hernández et al., 2014; 22%, Greenley et al., 2012; 33%, O’Bryhim et al., 2012). In most studies, a large number of potential loci are discarded because of amplification problems, which is particularly frequent among bivalves (Selkoe & Toonen, 2006). Mollusk genomes have been found to possess a high frequency of repetitive elements that may interfere if they appear in the flanking regions of microsatellite loci, resulting in multicyc copy PCR products (McInerney et al., 2011). Also, sometimes the use of the M13 tail for fluorescently labeling the forward primer may decrease the efficiency of PCR reactions (Guichoux et al., 2011). We did not find evidence of genotyping errors.

Most loci were characterized by moderate to high genetic variation, with an average of 11.4 alleles per locus (range = 5-22 alleles), heterozygosity estimates ranging between 0.429 and 0.818 (mean = 0.613) and PIC value between 0.488 and 0.939 (mean = 0.743). Three loci (Pgen3_7, Pgen4_1 and Pgen4_11) significantly deviated from HWE after the Dunn-Šidák correction (p < 0.006) due to heterozygote deficiencies, for which MICRO-CHECKER suggested the presence of null alleles. We found evidence of LD (p < 0.001) between Pgen4_1 and Pgen4_10. A similar range of alleles (4-23 per locus) were found in the congener *P. abbreviata* in microsatellite loci (n = 21) obtained from NGS (Ananchede et al., 2013) and similar results have been reported in others mollusk (An & Lee, 2012; Greenley et al., 2012). The quality of a marker can be determined by its degree of polymorphism. In this study, the expected heterozygosity values are in the optimal range (0.6-0.8) to provide a good resolution (Taberlet & Luikart, 1999). Moreover, the PIC values of all loci are higher than 0.25, so they are informative for linkage analysis. The deviation from HWE found in some loci could be due to population phenomena such as inbreeding, Wahlund effect, and selection or genotyping errors, such as null alleles or homoplasy (Selkoe & Toonen, 2006). However, since populations of *P. generosa* have shown no genetic differentiation along the northeast Pacific it is unlikely that a Wahlund effect is playing a part in the observed disequilibria (Suárez-Moo et al., 2016). On the other hand, MICRO-CHECKER analysis indicated the presence of null alleles in some locus. Null alleles have been found to be very common in bivalves because of mutations in the flanking regions (Becquet et al., 2009; Hedgecock et al., 2004). Even though population genetic studies require the use of independent unlinked loci, linkage may be useful for mapping studies (Xiao et al., 2012).
<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer sequence (5’-3’)</th>
<th>Tm</th>
<th>Motif</th>
<th>n</th>
<th>Na</th>
<th>Allelic range</th>
<th>Dye</th>
<th>$H_0$</th>
<th>$H_e$</th>
<th>$P_{HWE}$</th>
<th>PIC</th>
<th>GenBank accession</th>
</tr>
</thead>
</table>
| Pgen2_3   | F: GCCGTGATTGCRGGTGAT  
R: CAGGCATCGTCTGTAATGG  | 55.6 | (AT)$_8$ | 34   | 5   | 159-167      | FAM | 0.559 | 0.673 | 0.311     | 0.6   | MF668230         |
| Pgen3_4   | F: ACGGCCAAGAACGAAATGG  
R: TTGGTGAGGGTTGTTGCAG  | 55.6 | (ACG)$_10$ | 17   | 6   | 316-337      | PET | 0.529 | 0.617 | 0.223     | 0.559 | MF668231         |
| Pgen3_7   | F: GACAAACACCGCCTACACTG  
R: TACGAATGCAGTACCAAGC  | 66   | (AAC)$_{17}$ | 29   | 15  | 345-393      | FAM | 0.552 | 0.912 | 0.000*    | 0.888 | MF668232         |
| Pgen4_1   | F: GSGTGAATTCATTGGGTA  
R: ACCACCTGGACACTCTTA  | 62   | (ACAG)$_{14}$ | 24   | 22  | 314-474      | VIC | 0.667 | 0.962 | 0.000*    | 0.939 | MF668233         |
| Pgen4_3   | F: GTTTGCTTTGTCTTGCAG  
R: GGATCCCTGGAAAGTGTGGT  | 62   | (AAAC)$_6$ | 35   | 7   | 236-320      | PET | 0.429 | 0.548 | 0.141     | 0.488 | MF668234         |
| Pgen4_9   | F: GTCAATCCAGCCAAGCACAG  
R: GCGTGTAGCCTCAATAGC  | 55.6 | (AATC)$_8$ | 33   | 13  | 281-381      | PET | 0.818 | 0.891 | 0.045     | 0.866 | MF668235         |
| Pgen4_10  | F: AACCGCAGCAAAACAAGTC  
R: ATCTTGCCGTAGGAGGCCG  | 56.6 | (ACG)$_6$ | 28   | 13  | 345-409      | VIC | 0.724 | 0.823 | 0.039     | 0.786 | MF668236         |
| Pgen4_11  | F: AAGTCAACCAGGATGTCAC  
R: CCATAAAGGGTCTACACGCG  | 66   | (ATCC)$_8$ | 32   | 10  | 242-278      | NED | 0.625 | 0.848 | 0.002*    | 0.816 | MF668237         |

Abbreviations: Tm (°C): annealing temperature, n: sample size; Na: number of alleles; Dye: fluorescent dye; $H_0$: observed heterozygosity; $H_e$: expected heterozygosity; $P_{HWE}$: Hardy-Weinberg equilibrium test p-value (*) significant after Dunn-Sidak correction; PIC: polymorphism information content.
Nowadays, NGS is the best option to identify a high number of microsatellite loci in non-model species because it is cheaper and faster than traditional methods (Castoe et al., 2010; Ekblom & Galindo, 2011). These technologies are enabling more extensive and robust genetic studies in a great variety of taxa (Castoe et al., 2012; Huang et al., 2015; Mira et al., 2014). In this study we developed a set of new polymorphic microsatellites in P. generosa. These markers will useful in genetic studies applicable to conservation and management fisheries and aquaculture activities.

ACKNOWLEDGMENTS

Financial support was provided by the COFUPRO grant RG-ACES-2013-031 to ARO. We thank Grupo Marítimo Miramar for providing live specimens. The first author benefited from a graduate fellowship CONACYT to support her M. Sc. program in Marine Ecology at Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE).

REFERENCES


Hedgecock, D., G. Li, S. Hubert, K. Buolm & V. Ries. 2004. Widespread null alleles and poor cross-species amplification of microsatellite DNA...


