© Indian Academy of Sciences

ONLINE RESOURCES

Identification of sixteen single-nucleotide polymorphism markers in the pearl oyster, *Pinctada fucata*, for population genetic structure analysis

XIANDE HUANG[1], SHANZENG WU[1,2], YUNYAN GUAN[1], YAOGUO LI[1,2] and MAOXIAN HE[1]*

[1]*Key Laboratory of Tropical Marine Bio-resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, People's Republic of China*
[2]*University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China*

[Huang X., Wu S., Guan Y., Li Y. and He M. 2014 Development of sixteen single nucleotide polymorphism markers in the pearl oyster, *Pinctada fucata* for population genetic structure analysis. *J. Genet.* **93**, e1–e4. Online only: http://www.ias.ac.in/jgenet/OnlineResources/93/e1.pdf]

Introduction

The pearl oyster, *Pinctada fucata*, a marine bivalve belonging to the family Pteriidae, is the primary species cultured for marine pearls in China and Japan (Zhang 2002). To increase the yield, successful techniques for artificial breeding have been established in China (Zhang 2002). However, with increasing acreage for artificial farming, several issues such as declining genetic diversity in culture are being taken into consideration in the modern pearl oyster farming industry. Further study on these topics would require the available genetic tools such as molecular markers (Liu and Cordes 2004).

Single-nucleotide polymorphisms (SNPs) are nucleotide variations in the DNA sequence of individuals. As the most abundant molecular markers in the genome, they are sequence-tagged markers with codominant inheritance, and thus are ideal for population genetic studies (Morin *et al.* 2004). Therefore, in recent years, SNP have been characterized and applied in many aquaculture species such as *Salmo salar* (Hayes *et al.* 2007), *Crassostrea virginica* (Zhang and Guo 2010) and *Chlamys farreri* (Li *et al.* 2013). However, to our knowledge, although the draft genome of *P. fucata* has been established which could provide a platform for the identification of selection markers (Takeuchi *et al.* 2012), there are no SNP loci which have been identified for the pearl oyster, *P. fucata*.

At present, there are many SNP genotyping systems (Kim and Misra 2007). For our research focussing on developing a limited number of SNP markers for population studies, fragment length discrepant allele specific PCR (FLDAS-PCR), an improved allele specific PCR (AS-PCR) (Newton *et al.* 1989), is more flexible with a reasonable efficiency and expense (Huang *et al.* 2005; Gaudet *et al.* 2007), and has also been applied the identification of SNP markers in mollusks, such as *Meretrix meretrix* (Li *et al.* 2010), *Crassostrea gigas* (Wang *et al.* 2010) and *Mytilus galloprovincialis* (Li *et al.* 2011). In this method, for each SNP locus, the primer set includes a reverse common primer (Pub) and two forward allele-specific primers (As1 and As2) with the 3′ terminal base of each specific primer matching one of the SNP alleles. The primers are designed to amplify a fragment ranging from 100 to 220 bp. Two GC-rich tails of different length, six bases for one primer and 14 for the second one, are attached to each of the two forward allele-specific primers to obtain a difference of 8 bp between the amplification products. A destabilizing mismatch within the four bases of the 3′ end is also added to improve the allele specificity.

In this study, we chose 57 SNP candidates and developed 16 biallelic SNP markers in the pearl oyster, *Pinctada fucata*. These are the first SNP markers identified for analysing population genetic structure in *P. fucata* and will provide a useful complement to the currently available genetic markers.

Materials and methods

The UniGene sequences which were obtained from high-throughput sequencing of cDNA library constructed by our research group were screened for SNP using the software SOAPsnp (short oligonucleotide alignment program:

*For correspondence. E-mail: hmx@scsio.ac.cn.

Keywords. single nucleotide polymorphisms; FLDAS-PCR; *Pinctada fucata*.

Li *et al*. 2008, Li R. H. *et al*. 2009; Li R. Q. *et. al* 2009; Huang *et al*. 2013). Each SNP-containing sequence was inspected for the suitability for primer design. All the SNP-containing sequences that contained sufficient flanking sequences (no less than 100 bp) and of good quality (no unknown bases) were identified as candidate SNP markers.

In this study, FLDAS-PCR was used for SNP genotyping, two allele-specific primers with different GC tails and a common reverse primer were designed by PRIMER 5.0 (Biosoft International, Palo Alto, USA) with criteria described by Wang *et al*. (2005). For the polymorphism analysis of each locus, 40 individuals were sampled randomly from a wild

Table 1. Characterization of 16 SNP loci in *Pinctada fucata* from 40 wild individuals.

Locus	Primer sequences (5′–3′)	T_a (°C)	Allele and product size (bp)	Minor allele and frequency	H_o	H_e	*P* value
PM5	As1: *GCGGGCAGGGCGGC*TGTGACTGCAGTGCATTAGGGT As2: *GCGGGC*TGTGACTGCAGTGCATTATTGC Pub: GGTCGGTGTATGGTGTGGTGAT	60	T 207 C 199	T 0.2949	0.3846	0.4212	0.5843
PM7	As1: *GCGGGCAGGGCGGC*TGACAGCTTATCAAGTCCCACT As2: *GCGGGC*TGACAGCTTATCAAGTCCGTCA Pub: CAAGACAATCACCTACACCTCC	53	T 193 A 185	A 0.2625	0.3250	0.3921	0.2825
PM8	As1: *GCGGGCAGGGCGGC*CGTACTAGATCTGCGGAGGTAT As2: *GCGGGC*CGTACTAGATCTGCGGAGCAAC Pub: GAAGATATGAATGTCACCAGGG	56	T 217 C 209	T 0.2051	0.2564	0.3303	0.1762
PM16	As1: *GCGGGCAGGGCGGC*TCGCCCATCATGTGACAGAAGT As2: *GCGGGC*TCGCCCATCATGTGACAGAAGG Pub: ATTCACGAAGCTGACGACAACG	60	T 163 G 155	T 0.3026	0.3947	0.4277	0.6314
PM19	As1: *GCGGGCAGGGCGGC*ATAGACGGGGTTTTCCTTGAGG As2: *GCGGGC*ATAGACGGGGTTTTCCTTAGGA Pub: GCCTTCTAAATGGAACTTCGCT	60	G 129 A 121	G 0.4375	0.3750	0.4984	0.1119
PM20	As1: *GCGGGCAGGGCGGC*CCCAACTATTCCACCAGGCATT As2: *GCGGGC*CCCAACTATTCCACCAGGGTTC Pub: ACGGCAATCTTTTTCGTCACTT	60	T 162 C 154	T 0.1842	0.2105	0.3046	0.0727
PM21	As1: *GCGGGCAGGGCGGC*TTTGGACGAAGACCAACATCAT As2: *GCGGGC*TTTGGACGAAGACCAACAGAAC Pub: AGAAAAATGAGAGACAACGGCT	56	T 201 C 193	T 0.3077	0.4615	0.4316	0.6556
PM23	As1: *GCGGGCAGGGCGGC*GATGTGTGATGAAGTAATTGCC As2: *GCGGGC*GATGTGTGATGAAGTAATGGCT Pub: GTATACAGCTTATACCGGCCTT	56	C 127 T 119	T 0.1875	0.2750	0.3085	0.4971
PM26	As1: *GCGGGCAGGGCGGC*AGCTAACTCTGGATCCTCCTTG As2: *GCGGGC*AGCTAACTCTGGATCCTCGGTA Pub: CCAGTTGTCACCTGGTAACTGT	56	G 176 A 168	A 0.1923	0.1795	0.3147	0.0122*
PM27	As1: *GCGGGCAGGGCGGC*ACCAATAAATACTGCTTCAGCC As2: *GCGGGC*ACCAATAAATACTGCTTCGGCT Pub: TTTAATTGATGGGTTGGATACC	56	C 179 T 171	C 0.2375	0.4750	0.3668	0.0169*
PM28	As1: *GCGGGCAGGGCGGC*TACATTTGTTCCAACAGGCGTT As2: *GCGGGC*TACATTTGTTCCAACAGGGCTG Pub: TAGGGTTATCTATGATCGGCGC	60	T 164 G 156	T 0.0641	0.1282	0.1215	0.6005
PM32	As1: *GCGGGCAGGGCGGC*GGAAATTGGTACAAAAAGTGCA As2: *GCGGGC*GGAAATTGGTACAAAAAGGCCT Pub: CGAACGTGCAAATAACTTCAAG	56	A 205 T 197	T 0.3250	0.4500	0.4443	0.9342
PM33	As1: *GCGGGCAGGGCGGC*TACGACGTCTGTGAACGGTCTC As2: *GCGGGC*TACGACGTCTGTGAACGGAGTT Pub: TCTCGTGAGTGAATTCTGAGGC	59	C 157 T 149	T 0.4359	0.3590	0.4982	0.0762
PM43	As1: *GCGGGCAGGGCGGC*TCCTAGATTTTGCGCAATGTCA As2: *GCGGGC*TCCTAGATTTTGCGCAATATCG Pub: CTTTGTTTGGGAAGCATCAACC	59	A 173 G 165	A 0.1974	0.3947	0.3211	0.0621
PM44	As1: *GCGGGCAGGGCGGC*ATTCCCGCGTTATTCATATAAC As2: *GCGGGC*ATTCCCGCGTTATTCATAATAG Pub: ATCAATTAGAGCCCAAGACAGC	58	C 218 G 211	C 0.2375	0.3750	0.3668	0.8839
PM50	As1: *GCGGGCAGGGCGGC*GACAGCTGCTATACGACTGGAT As2: *GCGGGC*GACAGCTGCTATACGACTGGAC Pub: AACACATTGACTGGGGTTTGAT	56	T 210 C 202	T 0.3462	0.4872	0.4585	0.6903

T_a, annealing temperature; H_o and H_e, observed and expected heterozygosity; *P* value, *P*-value of Hardy–Weinberg equilibrium; *significant deviation from HWE; As1 and As2, two forward allele-specific primers; Pub, reverse common primer; The GC tails are italic bases and the mismatches are underlined.

Table 2. Functional annotation of 16 polymorphic SNP loci in *P. fucata*.

Locus	SNP type	Base location	Amino acid	Putative function	*E* value
PM5	T/C	–	–	Platelet glycoprotein V	$4.00E^{-37}$
PM7	T/A	TCT—TCA	Ser	Vasculin-like protein 1	$6.00E^{-65}$
PM8	T/C	TAT—TAC	Tyr	Unknown	**
PM16	T/G	GAC—GCC	Asp—Ala	Tripartite motif-containing protein 3	$1.00E^{-26}$
PM19	G/A	GGC—GGT	Gly	Selenium-dependent glutathione peroxidase	$1.00E^{-46}$
PM20	T/C	TAT—CAT	Tyr—His	Thrombospondin-1	$3.00E^{-24}$
PM21	T/C	CAT—CAC	His	Chitin synthase	$1.00E^{-109}$
PM23	C/T	GCC—GCT	Ala	T-complex protein 1 subunit gamma	0.00
PM26	G/A	TTG—TTA	Leu	Antisecretory factor-like protein	0.00
PM27	C/T	GCC—GCT	Ala	Cartilage matrix protein	$6.00E^{-11}$
PM28	T/G	GTT—GTC	Val	Cartilage matrix protein	$6.00E^{-11}$
PM32	A/T	CCA—CCT	Pro	Kyphoscoliosis peptidase	$1.00E^{-95}$
PM33	C/T	GTC—GTT	Val	Macoilin	$3.00E^{-105}$
PM43	A/G	GAT—GAG	Asp—Glu	Transcription factor LBX1	$4.00E^{-85}$
PM44	C/G	AGT—ACT	Ser—Thr	Unknown	**
PM50	T/C	GAT—GAC	Asp	Kinesin-like protein KIF21A	$2.00E^{-137}$

The underlined base, the base substitution in the corresponding sequence; –, indicates SNPs in noncoding regions; ∗∗, *E* value >$1.00E^{-5}$.

population in Shenzhen, Guangdong, China. Genomic DNA was extracted from adductor muscle using the E.Z.N.A Mollusc DNA Kit (Omega Bio-Tek, Norcross, USA).

Primers were first examined for PCR in eight wild individuals. PCR was performed in 20 μL reaction volume containing 12.3 μL H_2O miliQ, 2 μL PCR buffer (Mg^{2+} Plus) 10×, 1.6 μL dNTP 2.5 mM, 0.8 μL of each primer 10 μM, 0.2 μL *Taq* DNA polymerase (TaKaRa, Dalian, China) 5 U/μL and 1.5 μL of genomic DNA 20 ng/μL. All the candidates were amplified under the following conditions: 4 min denaturation at 94°C, 30 cycles of 30 s at 94°C, 30°C at 10 gradient annealing temperatures (52 to 62°C), and 30 s at 72°C, then a final extension at 72°C for 8 min. The products were separated with electrophoresis on nondenaturing polyacrylamide gels (10%) and visualized with silver staining. Once the corresponding PCR fragments were validated to be stable, specific polymorphic in eight wild individuals, polymorphism assessment in a wild population comprising 40 individuals was determined. The optimal annealing temperature was selected and other PCR conditions were followed above.

The size of allele was estimated using LabImage ver. 3.4 (http://labimage.software.informer.com/). Genetic variation in terms of the minor allele frequency, observed (H_o) and expected (H_e) heterozygosity and the probability of a deviation from Hardy–Weinberg equilibrium (HWE) were calculated using PopGene ver. 1.32 (Yeh *et al.* 1999). SNP-containing sequences were annotated by BLASTx in NCBI (http://www.ncbi.nlm.nih.gov/) and significant homology was accepted based on a cut-off *E* value of 1.0×10^{-5}.

Results and discussion

Through bioinformatic mining, 57 SNPs occurring in 56 sequences were identified as candidates SNP markers, PM27 and PM28 are located in one sequence with a spacing of 176 bp. These 57 SNPs could be classified into four motifs by base substitution type: 29 (50.9%) were A/G (C/T), 15 (26.3%) were A/C (G/T), nine (15.8%) were A/T, and only four (7.0%) were C/G. Besides, the classification result showed that C/T transition was the most frequent motif accounting for 36.8% (21) totally, and C/G transversion was the least common variation, which is same as *S. salar* and *M. galloprovincialis* (Hayes *et al.* 2007; Li *et al.* 2011).

Fifty-seven primer groups were designed to identify SNP markers. PCR results showed that 37 primer groups were successfully amplified, 16 (28.1%) of them generated clear and specific products consistently, showing biallelic polymorphisms in 40 samples from wild population. The efficiency in *P. fucata* is slightly higher than that in *C. gigas* (Wang *et al.* 2010). All the 16 SNP loci had two alleles of 119 to 219 bp and the minor allele frequency ranged from 0.0642 to 0.4375. The observed (H_o) and expected (H_e) heterozygosities ranged from 0.1282 to 0.4872 and from 0.1215 to 0.4984, respectively. Significant deviation from HWE ($P < 0.05$) was observed at two loci (PM26 and PM27) (table 1).

Using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), 15 genotyped SNP loci were found within ORF of the corresponding sequences, and only one locus, PM5, was located in untranslated regions (UTRs). There were 11 SNP loci identified as synonymous mutations and four SNP loci (PM16, PM20, PM43 and PM44) resulting in nonsynonymous amino acid substitutions (table 2). Li *et al.* (2013) found eight nonsynonymous mutation loci from 33 SNP loci in *C. farreri*. However, the vast majority of SNP markers reported in other mollusks, such as *Argopecten irradians irradians*, *C. gigas*, *M. galloprovincialis* are synonymous mutations (Li R. H. *et al*. 2009; Li R. Q. *et al*. 2009). Compared to 15 SNP-containing

sequences in the public database using BLASTx (NCBI), there were 13 SNP-containing sequences which were significantly ($E \leq 1 \times 10^{-5}$) homologous to genes with known functions. The putative functions of these SNP-containing sequences were related, respectively, to selenium-dependent glutathione peroxidase, chitin synthase, transcription factor LBX1, and so on (table 2).

In conclusion, this study reports the first SNP markers *P. fucata*. These 16 SNP markers will provide a useful tool for population genetics and evolutionary analysis in *P. fucata*.

Acknowledgements

This study was financially supported by the National Science and technology programme of China (2012AA10A410), the Funds of Knowledge Innovation programme of Chinese Academy of Sciences (ZCX2-EW-Q21), and Marine fishery science and technology promotion programme of Guangdong Province, China (A201201A05).

References

Gaudet M., Fara A. G., Sabatti M., Kuzminsky E. and Mugnozza G. S. 2007 Single-reaction for SNP genotyping on agarose gel by allele-specific PCR in black poplar (*Populus nigra*). *Plant Mol. Biol. Rep.* **25**, 1–9.

Hayes B., Laerdahl J. K., Lien S., Moen T. and Berg P. 2007 An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. *Aquaculture* **265**, 82–90.

Huang D. X., Yang Q. E. and Zhao G. S. 2005 A simple and rapid modified new method for SNP typing by fragment length discrepant allele specific PCR. *J. Forensic Med.* **21**, 11–14.

Huang X. D., Zhao M., Liu W. G., Guan Y. Y., Shi Y., Wang Q. *et al.* 2013 Gigabase-scale transcriptome analysis on four species of pearl oysters. *Mar. Biotechnol.* **15**, 253–264.

Kim S. and Misra A. 2007 SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.* **9**, 289–320.

Li H., Zhu D., Gao X., Li Y. F, Wang J. and He C. 2010 Mining single nucleotide polymorphisms from EST data of hard clam *Meretrix meretrix*. *Conserv. Genet. Resour.* **2**, 69–72.

Li H. J., Liang Y., Xing K., Su H., Gao X. G., Sui L. J. and He C. B. 2011 Development of single nucleotide polymorphism markers for blue mussel (*Mytilus galloprovincialis*) using expressed sequence tags. *J. Fish. China* **35**, 348–355.

Li J. Q., Bao Z. M., Li L. and Hu X. L. 2013 Development and characterization of EST-SNP in *Chlamys farreri*. *Period. Ocean Univ. China* **43**, 56–63.

Li R. H., Li Q. and Kong L. F. 2009 Characterization of expressed sequence tag-derived single-nucleotide polymorphisms in the bay scallop *Argopecten irradians irradians*. *Jpn. Soc. Fish. Sci.* **75**, 1389–1400.

Li R. Q., Li Y. R., Kristiansen K. and Wang J. 2008 SOAP: Short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714.

Li R. Q., Li Y. R., Fang X. D., Yang H. M., Wang J., Kristiansen K. and Wang J. 2009 SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **19**, 1124–1132.

Liu Z. J. and Cordes J. F. 2004 DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1–37.

Morin P. A., Luikart G. and Wayne R. K. 2004 SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* **19**, 208–216.

Newton C. R., Graham A., Heptinstall L. E., Powell S. J., Summers C., Kalsheker N., Smith J. C. and Markham A. F. 1989 Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503–2516.

Takeuchi T., Kawashima T., Koyanagi R., Gyoja F., Tanaka M., Ikuta T. *et al.* 2012 Draft genome of the pearl oyster, *Pinctada fucata*: a platform for understanding bivalve biology. *DNA Res.* **19**, 117–130.

Wang J., Chuang K., Ahluwalia M., Patel S., Umblas N., Mirel D. *et al.* 2005 High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. *Biotechniques* **39**, 885–892.

Wang S. Z., Li L., Qi H. G. and Zhang G. F. 2010 Development of 17 single nucleotide polymorphisms (SNPs) in *Crassostrea gigas* searched from EST database. *Oceanol. Limnol. Sin.* **41**, 274–281.

Yeh F. C., Yang R. C., Boyle T., Ye Z. H. and Mao Z. X. 1999 POPGENE version 1.32, the user friendly software for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

Zhang L. 2002 The development of international pearl industry and the counter measures taken by China to speed up Chinese pearl industry. *Mar. Sci.* **26**, 10–13.

Zhang L. S. and Guo X. M. 2010 Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. *Aquaculture* **302**, 124–129.

Received 16 June 2013, in revised form 4 August 2013; accepted 24 September 2013
Published on the Web: 3 March 2014