ONLINE RESOURCES

Characterization of 108 novel expressed sequence tag-derived single nucleotide polymorphism markers in the blood clam *Tegillarca granosa* using a transcriptome database

SHUANGSHUANG TENG[1,2,3], XINGGUAN LIN[1,2,3], JUN FANG[1,2,3], XUELIANG CHAI[1,2,3] and GUOQIANG XIAO[1,2,3]*

[1] *Zhejiang Mariculture Research Institute, Wenzhou 325005, China*

[2] *Engineering Research Center for Marine Bivalves, Chinese Academy of Fishery Sciences, Wenzhou 325005, China*

[3] *Zhejiang Key Laboratory of Exploitation and Preservation of Coastal Bio-Resource, Wenzhou 325005, China*

Keywords. *Tegillarca granosa*; polymorphic markers; non-synonymous substitution; high-resolution melting; SNPs

Running title: Characterization of 108 EST-SNPs in *Tegillarca granosa*

*For correspondence. E-mail: xiaogq1978@163.com.

Introduction

Tegillarca granosa is one of the most economically important bivalve species in China and southeast Asia with wide distribution in Indo-Pacific tropical to temperate estuaries (Wu *et al*. 2004; Funge-Smith *et al*. 2012). According to statistical data published in the Chinese Fishery Statistical Yearbook (Dong 2012; Dong 2013; Dong 2014; Yuan and Zhao 2015; Yuan and Zhao 2016), the annual yield of *T. granosa* varied between 140,000 and 180,000 tons,

constituting approximately 50% of Chinese clam production between 2010 - 2014. The yield of *T. granosa* is typically very large and is significantly higher than that of *Scapharca subcrenata* and *S. broughtonii*. Among the family Arcidae, these three clams represent shellfish of immense economical value. However, the wild stocks of *T. granosa* have been declining dramatically for last decade due to over-exploitation and the deterioration of environmental conditions in China. Given its important economic value and its descending tendency, therefore, it has become very important to develop ways of genetically conserving wild resources.

With recent advances in technological developments, especially with regard to next generation sequencing, single nucleotide polymorphism markers (SNPs) are becoming more and more popular in aquaculture research. SNPs are attractive for many reasons, such as the most abundant genetic variation, high frequencies in genomes (Vera *et al*. 2011), amenable for high throughput genotyping, easy standardization among laboratories and lower genotyping error rates (Garvin *et al*. 2010). In addition, SNPs confer several other key advantages. Firstly, it is possible to construct high-density genetic linkage maps with extremely high SNP density, due to their abundance in the genome. Secondly, it becomes possible to target related genes, thus creating a significant benefit in the genetic analysis of quantitative trait loci (QTL), if expressed sequence tag (EST)-SNPs deriving from a transcriptome database are used to QTL study (Muchero *et al*. 2011). EST-derived SNPs have been extensively reported in several marine molluscs, such as *Crassostrea gigas* (Zhong *et al*. 2013), *Hyriopsis cumingii* (Yin *et al*. 2015), *Meretrix meretrix* (Jing *et al*. 2015), and *Chlamys farreri* (Li *et al*. 2013). However, the number of available SNPs, which was 90, remains limited for *T. granosa* (Dong *et al*. 2014). Thus, a greater number of SNP markers are required for *T. granosa* in order to construct high-density genetic linkage maps, locate definitely quantitative trait loci, carry out genome-based genetic surveys in molecular breeding programs, and perform marker-assisted selection.

Over recent years, high resolution melting (HRM) has become a very promising technique of SNP detection for high throughput, fast, and cost-effective (Li *et al*. 2012). In this technique, SNPs are identified by changes in the melting curves of PCR products, and by different alleles exhibiting different impacts upon annealing temperature. This method has

been applied for the rapid detection and typing of SNPs which are valuable tools for the investigation of genes associated with the regulation of certain economically-important characteristics, such as growth, reproduction, nutritional content, and resistance to disease in aquatic organisms (Sepulveda *et al*. 2012; Li *et al*. 2013).

In the present study, we developed and characterized 108 EST-SNP markers from a *T. granosa* transcriptome database using HRM genotyping methodology. These markers represent critical tools for genetic studies in this highly valuable species of clam and make a significant contribution to the ongoing development of conservation and management programs.

Materials and methods

Sample collection and DNA extraction

In order to screen polymorphic SNPs, we collected 30 wild *T. granosa* from YueQing Gulf of Zhejiang Province, China. For each individual, genomic DNA was extracted from the adductor muscle by standard proteinase K digestion, phenol-chloroform extraction and ethanol precipitation, as described by Li *et al*. (2002).

Data mining for SNP markers, primer design and amplification

In our previous studies, we sequenced the *T. granosa* transcriptome using 454 high-throughput sequencing technology. Raw reads from this transcriptome were assembled and clustered into contigs which contained at least 4 reads and these were then used to detect putative SNPs using the QualitySNP program (Tang *et al*. 2006). A single-base mutation existing in two or more reads were selected for further analysis. At least 50 bases stood at either end of putative SNPs which aided in the design of specific primers. Primers were designed with Primer Premier 5.0 software (http://www.premierbiosoft.com/primerdesign/) to have an annealing temperature of 55 - 60 °C and a product length of 100~200 bp. Ten *T. granosa* individuals were chosen randomly, and DNA templates from these ten individuals were mixed together in a balanced manner in order to validate whether the primers were designed appropriately.

PCR amplifications were performed in a total reaction of 20 μL containing 1 U r*Taq* DNA polymerase (Takara, Japan), 1×PCR buffer, 0.2 mM dNTP, 0.25 μM of each primer set, and

approximately 100ng of DNA template. PCR amplification was then carried out using the following cycling conditions: initial denaturation at 94 ºC for 3 min, then 35 cycles of denaturation at 94 °C for 30 s, annealing temperature for 30 s, and extension at 72 °C for 45 s, with a final extension at 72 °C for 7 min. Amplification products were visualized by 8% polyacrylamide gel electrophoresis. Primers which were successfully able to amplify the expected fragment length of DNA were chosen for further genotyping.

Characterization of polymorphic SNPs

To evaluate polymorphisms in our selected markers, the 30 collected individual of *T. granosa* were employed for genotyping using the 7500 Fast Real-Time PCR System (Applied Biosystems, Waltham, MA, USA). The PCR reaction mixture was composed of 10 μL Melt Doctor HRM Master Mix (Applied Biosystems), 0.25 μM of each primer sets, approximately 100 ng of DNA template, and RNase-free water to a final volume of 20 μL. Amplifications were performed as described above, and then melting curves were prepared under the following conditions: denaturation for 10 s at 95 °C and annealing for 1 min, then high resolution melting for 15 s at 95 °C and annealing for 15 s. Data were recorded and analyzed using HRM system software (Applied Biosystems).

Statistical analysis

Minimun allele frequency (MAF), expected and observed heterozygosity (H_e and H_o), linkage equilibrium and Hardy-Weinberg equilibrium (HWE) were tested using Popgene32 software (Yeh & Boyle 1997). Significant levels were calculated per locus using Bonferroni method (Rice 1989). The sequences containing SNPs were annotated using BLASTx against non-redundant GenBank and Swiss-Prot databases. The sequence homology was accepted based on a cut-off E-value of 1.0×10^{-5}. The informative strand and reading frame were identified by using the sequence with highest homology. The Open Reading Frame Finder (https://www.ncbi.nlm.nih.gov/orffinder/) was used to determine whether SNPs were synonymous, nonsynonymous, or from untranslated regions (UTRs).

Results and discussion

In the present study, our *T. granosa* transcriptome database, with thousands of polymorphic sequences, was utilized for SNP screening and validation. Sequences were assembled and

clustered into 13,290 contigs, of which 3258 contigs contained 13,193 SNPs. A total of 3982 putative SNPs were detected within the 1310 contigs containing at least 4 reads. The number of SNPs per contig ranged from 1 to 24 with an average of 3.0. These mutations included 2265 transitions (A/G 1136 and T/C 1129) and 1717 transversions (A/C 348, A/T 847, G/T 345, G/C 177).

In total, 183 putative SNPs with conserved flanking sequences were selected for validation. Only one SNP was analyzed at each contig. Of the 183 primer sets designed to amplify specific DNA fragments, 144 sets successfully amplified fragments of the expected length, while the other 39 primer sets failed. Primer site polymorphism or interference of amplicon by intron(s) may be responsible for miss-amplification (Li *et al*. 2011).

Among the remaining 144 candidate SNPs yielding products with expected sizes, 14 loci could not be successfully distinguished. Among the 30 wild blood clam investigated, 108 loci (59.0%) represented polymorphic SNPs (Table 1), 18 loci were homozygous in all individuals, and four loci were heterozygous in all individuals. These four loci might represent sequence variation between paralogs or technical genotyping errors, rather than orthologous alleles (Moen *et al*. 2008; Hubert *et al*. 2010; Vera *et al*. 2011), The higher validation rate, compared with *C. farreri* (33.7%) (Li *et al*. 2013), *C. gigas* (24.0%) (Zhong *et al*. 2013), *M. meretrix* (22.3%) (Jing *et al*. 2015), may be the consequence of effective primer design. Primers were designed to locate SNPs in the middle of the PCR products which were approximately 100 bases in length. As the amplicon size decreased, we obtained better differentiation using the HRM method. Wang et al (2008) identified two significant factors to consider in the validation of EST-derived SNPs: contig size (the number of sequences in the contig) and minimun allele sequence frequency. We observed that the larger the contigs, the greater the validation rate. Validation rate was reasonably high when the minimun allele sequence was represented at least twice within contigs which contained four or more EST sequences.

Further evaluation in the *T. granosa* population showed that, a total of 67 transitions (A/G 28 and C/T 39), 41 transversions (A/C 11, A/T 16, C/G 2, G/T 12) were detected. The observed and expected heterozygosity ranged from 0.100 to 0.900 and 0.097 to 0.514, respectively, with respective means of 0.455 and 0.436. Minimum allele frequency ranged from 0.050 to 0.500, while HWE *p*-values varied from 0.0004 to 0.9445. Following

Bonferroni correction for multiple comparisons, only one SNP locus was found to significantly deviate from HWE and no significant linkage disequilibrium was detected between the 108 loci tested.

Of the 108 polymorphic SNPs, 27 (25%) could be annotated and were found to encode important proteins such as transcriptional regulator, intracellular protein transporter, and post-translational modification. Twelve SNPs were located in the untranslated region (UTR) and fifteen in the coding region. Seven of these 15 SNPs were synonymous and eight non-synonymous. These non-synonymous SNPs may change protein structure and function in the correspondent to phenotypic and trait changes.

In conclusion, a total of 108 SNPs were developed using the HRM method for *T. granosa*. These markers will be useful for studying the population structure of this species, and in characterizing variation in natural and wild stocks, performing parentage identification and marker-assisted selection. Such analyses will provide important information relating to stock-specific abundance, and contribute to the conservation management of this species.

Acknowledgements

This work was supported by the National High Technology Research and Development Program of China (2012AA10A410), the National Project for Agricultural Technology System (CARS-48), Zhejiang science and Technology Plan Program (2017F30042), and the Seed Science and Technology Innovation Project of Wenzhou City (N20120017).

References

Dong J. H. 2012 *Chinese Fishery Statistical Yearbook*, pp. 33. Chinese Agriculture Press, Beijing.

Dong J. H. 2013 *Chinese Fishery Statistical Yearbook*, pp. 33. Chinese Agriculture Press, Beijing.

Dong J. H. 2014 *Chinese Fishery Statistical Yearbook*, pp. 33. Chinese Agriculture Press, Beijing.

Dong Y. H., Shi S. F., Bao Y.B., Yao H. H., Jing Y. and Lin Z. H. 2014 Development of 90 EST-SNP markers in blood clam (*Tegillarca granosa*) using high resolution melting (HRM). *Conserv. Genet. Resour.* **7**, 309-314.

Funge-Smith S., Briggs M. and Miao W. M. 2012 *Regional overview of fisheries and aquaculture in Asia and the Pacific 2012*, pp. 114. RAP Publication, Bangkok.

Garvin M., Saitoh K. and Gharrett A. 2010 Application of single nucleotide polymorphisms to non-model species: a technical review. *Mol. Ecol. Resour.* **10,** 915-934.

Hubert S., Higgins B., Borza T. and Bowman S. 2010 Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genomics* **11**:191.

Jing Y., Dong Y. H., Yao H. H., He L. and Lin Z. H. 2015 Characterization of 87 EST-SNP markers in hard clam *Meretrix meretrix* using high-resolution melting analysis. *Conserv. Genet. Resour.* **7,** 811-817.

Li F. B., Niu B. L., Huang Y. P. and Meng Z. Q. 2012 Application of high-resolution DNA melting for genotyping in Lepidopteran non-model species: *Ostrinia furnacalis* (Crambidae). *Plos One* **7**, e29664.

Li J. Q., Bao Z. M., Li L., Wang X. J., Wang S. and Hu X. L. 2013 Development of 101 novel EST-derived single nucleotide polymorphism markers for Zhikong scallop (*Chlamys farreri*). *J. Ocean Univ. China* **12**, 403-412.

Li Q., Park C. and Kijima A. 2002 Isolation and characterization of microsatellite loci in the Pacific abalone, *Haliotis discus hannai*. *J. Shellfish Res.* **21**, 811-815.

Li Q., Wang Q. Z., Qi M. J., Ge J. L. and Cong R. H. 2011 Development, characterization, and inheritance of 113 novel EST-SSR markers in the Pacific oyster (*Crassostrea gigas*). *Genes Genom.* **33**, 313-316.

Moen T., Hayes B., Nilsen F., Delghandi M., Fhalestad K., Fevolden S. E. *et al*. 2008 Identification and characterization of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC genet.* **9**: 18.

Muchero W., Ehlers J. D., Close T. J. and Roberts P. A. 2011 Genic SNP markers and legume synteny reveal candidate genes underlying QTL for *Macrophomina phaseolina* resistance and maturity in cowpea (*Vigna unguiculata* (L) Walp). *BMC Genomics* **12**, 8.

Rice W. R. 1989 Analyzing tables of statistical tests. Evolution, 43(1): 223-225

Sepulveda D., Cardenas C., Carmona M. and Marshall S. H. 2012 Novel strategy to evaluate infectious salmon Anemia virus variants by high resolution melting. *Plos One* **7**, e37265.

Tang J., Vosman B., Voorrips R. E, van der Linden C G. and Leunissen J. A 2006 QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. *BMC bioinformatics* **7**, 438.

Vera M., Alvarez-Dios J. A., Millan A., Pardo B. G., Bouza C., Hermida M. *et al*. 2011 Validation of single nucleotide polymorphism (SNP) markers from an immune

expressed sequence tag (EST) turbot, *Scophthalmux maximus*, database. *Aquaculture* **313**, 31-41.

Wang S. L., Sha Z. X., Sonsregard T. S., Liu H., Xu P., Somridhivej B. *et al*. 2008 Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics* **9**, 450-460.

Wu H. X. and Chai X. L. 2004 The reproductive habit and growth character of bloody clam *Tegillarca granosa* in Yueqing bay. *Chin. J. Zool.* **39**, 47-50.

Yeh F. C. and Boyle T. J. B. 1997 Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg. J. Bot.* **129**, 157-163.

Yin H., Bai Z. Y. and Li J.L. 2015 Development of 56 novel EST-SNP markers in *Hyriopsis cumingii* based on transcriptome database. *Conserv. Genet. Resour.* **7**, 793-796.

Yuan X. C. and Zhao W. W. 2015 *Chinese Fishery Statistical Yearbook*, pp. 33. Chinese Agriculture Press, Beijing.

Yuan X. C. and Zhao W. W. 2016 *Chinese Fishery Statistical Yearbook*, pp. 33. Chinese Agriculture Press, Beijing.

Zhong X. X., Li Q., Yu H. and Kong L. F. 2013 Development and validation of single-nucleotide polymorphism markers in the Pacific oyster, *Crassostrea gigas*, using high-resolution melting analysis. *J. World aquacult. Soc.* **44**,

455-465.

Received 6 April 2017, final revised form 8 June 2017; accepted

13 June 2017

Table 1. Characterization of the 108 novel EST-SNP markers derived from *Tegillarca granosa*

Locus	Variants	SNP position (bp)	Primer sequence(5' -3')	T_a(℃)	H_o	H_e	MAF	HWE	Functional annotation	E-value	SNP location/effect
TgSNP_1	G/A	937	F: TTGTATACTTTGAAATGTTGACTGC	56.6	0.450	0.512	A=0.475	0.5809	not significant homology		Unkonwn
			R: GAAATAATGCTTGATGATTGTCTT								
TgSNP_2	C/T	97	F: TCAGCTATCTATGTCCTTAATGTGT	55.4	0.105	0.512	T=0.4737	0.0004	not significant homology		Unkonwn
			R: CATGAATAACCAGTGATACAATTTC								
TgSNP_3	G/T	264	F: TTTAATTTAATGTATCTGACATGCC	56.1	0.600	0.467	T=0.350	0.1875	not significant homology		Unkonwn
			R: CAACATGGTATCACATATGCAGT								
TgSNP_4	G/T	182	F: TTGATAACCAGTTCCAATTAGCTA	56.4	0.800	0.508	G=0.450	0.0082	not significant homology		Unkonwn
			R: TCTCTATTGCTGAATGTTATTTTCT								
TgSNP_5	A/G	154	F: CAAGTTGTGTGAATTTTATGGAAG	56.7	0.150	0.409	A=0.275	0.0032	not significant homology		Unkonwn
			R: AATTTGTTAGGATGAAATAAACATG								
TgSNP_6	A/C	313	F: TGATACTGCCACTGTTAACAAGG	57.5	0.700	0.513	A=0.500	0.0940	not significant homology		Unkonwn
			R: TGTTGTCAAACTTTTTAACTGCTG								
TgSNP_7	A/G	294	F: ATAACTCGTCAAGCTGAAAAGG	56.2	0.900	0.513	A=0.500	0.0005	not significant homology		Unkonwn
			R: CTTGATTTTGGGTTTGTTTGTT								
TgSNP_8	G/T	126	F: CCAAATTCTGCTAATCGTAAGTG	56.9	0.500	0.492	G=0.400	0.9427	28S ribosomal protein S12, mitochondrial	1.00E-31	UTR
			R: GTTATGTTCTCTGTAAATGGTGTCC								Ser(AGC) >Arg(CGC)
TgSNP_9	C/T	205	F: TTCTTTCCTTTCAACTTCATCTT	55.2	0.400	0.508	C=0.450	0.3302	Calreticulin	2.00E-46	Non-synonymous[1]
			R: AAAAATATGCCTCACAAAAAGC								Asn(AAT)>Asp(GAT)
TgSNP_10	C/T	70	F: GGTGACACAGAAATTCTCATGTATT	57.3	0.400	0.431	T=0.300	0.7405	60S ribosomal protein L15	2.00E-102	Synonymous[1]
			R: TGGAAACGCAGGAATACTCAG								Stop codon(TGA)>(TAA)
TgSNP_11	A/G	232	F: AGGTCAGATGTAGCCGATATTTA	55.9	0.800	0.508	A=0.450	0.0082	ribosomal protein rpl38	5.00E-24	UTR

			R: AGCAAGACCTGGAGGGAGT								
TgSNP_12	T/C	334	F: ATACCTATGTTGCCATTGACG	55.4	0.800	0.492	T=0.400	0.0041	not significant homology		Unkonwn
			R: CCTTCTCGAATTTTTCTAGTCATAT								
TgSNP_13	A/C	86	F: AATTGCCTTTTCACTATTATCATGT	57	0.650	0.481	A=0.375	0.1050	not significant homology		Unkonwn
			R: ACTACTGTCCAGCATCTTCATCT								
TgSNP_14	A/C	575	F: TTGTAAATCCGCAGTATCAGTG	55.9	0.350	0.512	A=0.475	0.1473	not significant homology		Unkonwn
			R: GGATCTTCCCATACAGGCTAA								
TgSNP_15	C/T	135	F: ATTATCGATGAAACGGATTACTT	55.2	0.750	0.501	C=0.425	0.0226	not significant homology		Unkonwn
			R: CATAAGGTCAAGATAAGCCAAAG								
TgSNP_16	C/T	57	F: CAAACAAGAAAAGTTTTGTATTGTC	55.8	0.278	0.475	T=0.361	0.0688	not significant homology		Unkonwn
			R: AAACCAAAACCGGATACATGTA								
TgSNP_17	A/C	1208	F: GTAATACTCAATTTGCTTTCGGA	56.5	0.550	0.409	A=0.275	0.1086	not significant homology		Unkonwn
			R: CAGAATGCTGCCATGGTATC								
TgSNP_18	G/T	598	F: GGCCAGGGGGACAATAAT	54.3	0.250	0.296	T=0.175	0.4562	La-related protein 1	2.00E-121	UTR
			R: ACGAAAGTCATTAAACAGAATACAGA								
TgSNP_19	A/G	453	F: TATGTGCCTTATCTAAAACAATCC	55.9	0.400	0.385	G=0.250	0.8517	not significant homology		Unkonwn
			R: ACCTTCCTGATAAAATAAAATGC								
TgSNP_20	A/T	288	F: GTAACAAAAGTGATAATTCAGACCA	55.7	0.600	0.431	T=0.300	0.0685	not significant homology		Unkonwn
			R: CCCTTTAATTGTATCATTTCCAT								
TgSNP_21	C/T	356	F: AAGTATAGGCGGATTTTGACC	55.8	0.550	0.501	C=0.425	0.6553	not significant homology		Unkonwn
			R: TGATTATGACCTTTGAGCTTACC								
TgSNP_22	A/T	206	F: TGGGACTTTTCAATGGGTAA	55.5	0.263	0.235	T=0.132	0.5593	not significant homology		Unkonwn
			R: AATCATATGTATTTTAAAGGAGCAA								
TgSNP_23	A/C	280	F: AAGTATATCCTCCATTTTGTTTTCT	55.7	0.500	0.492	C=0.400	0.9427	not significant homology		Unkonwn
			R: GCCAGAATGTTAGCTCAAAAAT								
TgSNP_24	C/G	407	F: TGGAAGTTAAATGGTCTAAAGCA	56.9	0.650	0.450	G=0.325	0.0399	not significant homology		Unkonwn

			R: TGTTAAAAAAGTTGCTACATAAAAGA								
TgSNP_25	C/G	670	F: GTTAATCACAATGCATTTAGCTG	55.3	0.400	0.328	G=0.200	0.3004	not significant homology		Unkonwn
			R: GGTTGTCTGTCTGTCAAGCCA								
TgSNP_26	A/G	529	F: TGTGTCCTAGGTATGCAAATCTTT	58.1	0.200	0.262	A=0.150	0.2541	not significant homology		Unkonwn
			R: TGAGACTAATGAGAAGCAGGAATC								
TgSNP_27	A/T	203	F: AGGGGGGATAGATTTACTTATACTA	55.3	0.500	0.385	T=0.250	0.1609	not significant homology		Unkonwn
			R: TACATGTACATGTTTACATGGGG								
TgSNP_28	C/T	502	F: AAAGGGCTTGATTTGAACG	55.4	0.300	0.492	T=0.400	0.0724	PREDICTED: uncharacterized protein LOC 105343024 [Crassostrea gigas]	4.00E-24	Non-synonymous[1]
			R: CAGAAGTGGAAGTGGGCAT								Gly(GGA)>Ary(AGA)
TgSNP_29	C/T	300	F: ATAACTGTGCTTTTCCCCACTATAA	59.1	0.500	0.508	T=0.450	0.9445	not significant homology		Unkonwn
			R: ACATCTTGAACATTCAAACTGTAAT								
TgSNP_30	A/G	542	F: TCAACGATAATAGCGAGGATAA	55.3	0.350	0.409	A=0.275	0.5023	not significant homology		Unkonwn
			R: GTTGCAACTACAATTAAGTTCAAGT								
TgSNP_31	C/T	107	F: CAGTCCTTTTATTCTCCAGATTGT	56.9	0.300	0.385	C=0.250	0.3039	60S ribosomal protein L5	2.00E-141	Non-synonymous[1]
			R: CGTGTGAAGATGTCCCGTG								Lys(AAA)>Glu(GAA)
TgSNP_32	A/G	447	F: TTTTAATTCTTTTGGTCAGTCTGT	55.4	0.550	0.501	G=0.425	0.6553	not significant homology		Unkonwn
			R: GAACAGTCAAAAGTATGCCAATC								
TgSNP_33	C/T	104	F: CTTGGTCGTGTCTAAATCGG	55.6	0.550	0.512	T=0.475	0.7301	60S ribosomal protein L12	8.00E-64	UTR
			R: CGGATCGAACTTAGGAGGC								
TgSNP_34	C/T	168	F: ACAATACTCTTCTGCATGTCTTTCT	56.7	0.350	0.481	C=0.375	0.2104	not significant homology		Unkonwn
			R: AAAAAATGAAAATTTAACAGCTTTA								
TgSNP_35	A/T	223	F: CTGTGTTTATGCCCGGTATG	56.5	0.400	0.328	T=0.200	0.3004	not significant homology		Unkonwn
			R: TCTCAGAAGTACGTGCTAACAATAT								
TgSNP_36	C/T	371	F: AGCCTACACCTGCCTGATACT	55.8	0.300	0.492	C=0.400	0.0724	not significant homology		Unkonwn
			R: TAATCAGGCCGGTCTAAGC								

TgSNP_37	C/T	99	F: CTGGGTTCATGTACTGAGCTCT R: GTCAAATTTAATGAAATGAATGAAT	56.2	0.250	0.358	C=0.225	0.1571	not significant homology		Unkonwn
TgSNP_38	G/T	249	F: GTCCTTCATTTCTCTGGTCTCTT R: TATCGGACGACTTGGACACT	56.3	0.350	0.481	G=0.375	0.2104	Nuclear factor erythroid 2-related factor 2	5.00E-25	Synonymous[1] Leu(CTA)>(CTC)
TgSNP_39	G/T	345	F: CGTGATGTACAATATTTACGTTTTA R: AACTGGGTAACTTAAACACAAACA	55	0.250	0.512	T=0.475	0.0190	hypothetical protein LOTGIDRAFT_209701[Lottia gigantea]	3.00E-19	UTR[1]
TgSNP_40	C/T	259	F: CCAGCAAACTCATTCCATCA R: TGAAGCACAACAAGTTTAATATAGC	56.5	0.550	0.501	T=0.425	0.6553	not significant homology		Unkonwn
TgSNP_41	A/G	192	F: ACAATCGAATAAGCTGGAATCT R: ACCATACTATGCCGCTTCTAA	55.5	0.300	0.492	G=0.400	0.0724	not significant homology		Unkonwn
TgSNP_42	C/T	401	F: GAGAATCTCTTATCTGCAGTCACTT R: AAGTCAGGTTGCATTCTAAAGTAA	56.3	0.150	0.409	C=0.275	0.0032	not significant homology		Unkonwn
TgSNP_43	A/C	561	F: CTTTACCTAAATGGGATATTTGAAA R: CAATGCTCCAAAAAGTTAAGATG	57.1	0.600	0.467	A=0.350	0.1875	not significant homology		Unkonwn
TgSNP_44	A/G	126	F: GAGAAAGAATTCAAGATGATGGTT R: GTTCTCCACTAATTACAACGCTC	56.7	0.450	0.409	A=0.275	0.6407	heat shock protein 20	6.00E-47	Synonymous Val(GTG)>(GTA)
TgSNP_45	A/T	306	F: CATTGTTGAAAAAAGGGTTAAAA R: ACCATCCTTGCATTATTATGTGTA	56.7	0.450	0.358	T=0.225	0.2252	not significant homology		Unkonwn
TgSNP_46	C/T	153	F: GACTCATGCTGACAAGAACCC R: AGTATAAGTAACCGTTTAAGATCCC	56.8	0.250	0.409	C=0.275	0.0705	not significant homology		Unkonwn
TgSNP_47	A/T	343	F: TGCATATATTTTTGTTGTGCTATTA R: TGCAAATTTGATCCTAAAATAATAA	55.6	0.200	0.185	T=0.100	0.6706	not significant homology		Unkonwn
TgSNP_48	A/G	113	F: AGGTCACCTGCTTGTCAAATC R: ATGGGTCTTAATAGGTAGATGGTC	56.6	0.700	0.508	A=0.450	0.0820	not significant homology		Unkonwn
TgSNP_49	A/G	227	F: TGAAGATGAACAGGTAGGACAGT	55.7	0.500	0.508	A=0.450	0.9445	not significant homology		Unkonwn

			R: CACCGTTTCCCTCCTCTAGA								
TgSNP_50	C/T	323	F: CGTTTTGTTGACGCCATCT R: GCAGACGGTTTAGCAATATTGA	57.7	0.400	0.467	C=0.350	0.5099	T-cell acute lymphocytic leukemia protein 1-like	3.00E-37	Non-synonymous[1] Gln(CAA)>Arg(CGA)
TgSNP_51	A/C	457	F: TGGGTACACTTAAAAAATTATTGTG R: AAAACTTTTCATTAGGAATCCATTA	56.3	0.650	0.512	A=0.475	0.2142	not significant homology		Unkonwn
TgSNP_52	A/G	130	F: ACCAGATTATGCAAAATTAAAAGTA R: TACTGTAAATTAGGAAATCTTCACG	55.6	0.550	0.512	G=0.475	0.7301	not significant homology		Unkonwn
TgSNP_53	C/T	232	F: CCTGTAACTTTCAGTGAAGCATT R: AAACAAGGTGTGAGAACAAAATT	56.2	0.556	0.514	C=0.500	0.7261	T-complex protein 1 subunit epsilon	1.00E-96	UTR
TgSNP_54	G/T	65	F: ACAGTGACAAGATAAGGTGGATAG R: AAAGCTTACAGATTAACGGCC	55.1	0.650	0.512	G=0.475	0.0024	not significant homology		Unkonwn
TgSNP_55	C/T	155	F: ACCAGTCCACACCGTCAAA R: GGAAAATTTTCAAATGAGTTTACAA	57.6	0.333	0.514	C=0.500	0.1245	elongation factor 1 gamma	1.00E-85	UTR[1]
TgSNP_56	A/G	716	F: ACATCAACACAAAAGAATCCG R: TTTTCTGAACAATGTGAACTGATA	55.2	0.250	0.409	G=0.275	0.0705	PREDICTED: tetraspain-9-like [Aplysia californica]	3.00E-23	UTR
TgSNP_57	G/T	54	F: AAGGCAGGCTTTCATCATTT R: ATCTCTGATCGATGTATTTTCTGTT	56.7	0.350	0.409	T=0.275	0.5023	not significant homology		Unkonwn
TgSNP_58	C/T	67	F: TTAAGTTGTTTTTCTTTTGCATG R: TAACACTCTGAAACTTCCAACG	55.4	0.650	0.512	T=0.475	0.2142	not significant homology		Unkonwn
TgSNP_59	A/T	786	F: CATCAACTTTACATAATTCTCCATC R: GTTGTTGGGGAGAGTAAGGTT	55.3	0.250	0.224	T=0.125	0.5720	not significant homology		Unkonwn
TgSNP_60	G/T	258	F: AGTGAAAAAACAACAAAGAAAATACA R: CCATTCACATGTTTGTTGAAATAT	55.8	0.450	0.409	G=0.275	0.6407	not significant homology		Unkonwn
TgSNP_61	C/T	114	F: ATCCATTACTTTTAGACTCCACTTG	56.5	0.300	0.328	T=0.200	0.6841	Elongation factor 1-beta	4.00E-16	Synonymous

			R: GGATGAAGAAGATGTGGCTGTT								Leu (CTA)>(TTA)
TgSNP_62	A/G	328	F: TTGAGAATAAGCACTTGAACTTAAA	56.2	0.700	0.492	G=0.400	0.0524	not significant homology		Unkonwn
			R: TTGTGTTGTCTACACTTATACTGGC								
TgSNP_63	C/T	415	F: ATTATATCTTGTGTCAAATGAACGA	56	0.550	0.501	T=0.425	0.6553	not significant homology		Unkonwn
			R: TTGTTTGATCCATTTCGCC								
TgSNP_64	A/G	293	F: AAATTTCCTCTGACCCTAACACA	57.7	0.550	0.512	G=0.475	0.7301	not significant homology		Unkonwn
			R: CAATTTTGTACACTGGTATACTCGA								
TgSNP_65	A/T	131	F: TGTGACAATCAAATCATGTACTTCA	57.7	0.100	0.097	T=0.050	0.8694	not significant homology		Unkonwn
			R: CATAGAAATACTTAGTATTTGGTGCATA								
TgSNP_66	A/G	373	F: GAAGTGTGAAACAGACAGTTATTGA	56.1	0.550	0.512	G=0.475	0.7301	not significant homology		Unkonwn
			R: TCCATGTGGTCAGTGAATATAAAT								
TgSNP_67	A/T	153	F: TTTAAGACACTTTTTGAAAATTGAA	56.2	0.250	0.224	T=0.125	0.5720	Actin-depolymerizing factor 2	1.00E-17	UTR[1]
			R: TGTTGGTGTAAATTATCAACAAATA								
TgSNP_68	A/T	469	F: TGATAAAATTCTATTCAAAAAACCC	57	0.150	0.142	T=0.075	0.7699	not significant homology		Unkonwn
			R: TCTTCTAGAAAATGCATTTACAGGT								
TgSNP_69	A/G	609	F: GTTAGATCTATATACACGGCTTTCA	55.5	0.500	0.467	G=0.350	0.7418	not significant homology		Unkonwn
			R: CAATAGATTTACAAACATGGATGG								
TgSNP_70	A/G	133	F: AAATACACCAGTACAAACAAAACAG	55.9	0.300	0.328	G=0.200	0.6841	not significant homology		Unkonwn
			R: CCTGTGAATTTATTATCTTTGTTTT								
TgSNP_71	C/T	582	F: TAAAGAAGGAAAGAAACCACAGT	55.2	0.650	0.512	C=0.475	0.2142	ribosomal protein S6	8.00E-125	Synonymous
			R: TTGAGTGCCAGCCTGTAGC								Val(GTT)>(GTC)
TgSNP_72	C/T	303	F: AGGCTGGCACCTAACAATCA	58	0.600	0.467	T=0.350	0.1875	not significant homology		Unkonwn
			R: CTTCTTTTCTGAAATAAACTTGTGG								
TgSNP_73	C/T	354	F: CTCAATTGTAAAAACCAAATAAGAA	55.8	0.550	0.481	T=0.375	0.5073	not significant homology		Unkonwn
			R: CTCGCTGATAATAATGACTCCA								
TgSNP_74	C/T	300	F: GTTCAGCAAGTGGCTCCTC	55.1	0.850	0.512	C=0.475	0.0024	not significant homology		Unkonwn

			R: CTTACCAGTCGAATAACAAACG								
TgSNP_75	A/T	329	F: CATTGACAAGAGTACAGGAAAAGA R: GCATCATTAACCATTCTGTCG	56.2	0.211	0.194	T=0.105	0.6608	heat shock protein 70	2.00E-122	Synonymous Thr(ACA)>(ACT)
TgSNP_76	C/T	351	F: TATGAACGCTTGGCTGCC R: GGCCTATCGATCCTTTTGA	57.7	0.450	0.512	C=0.475	0.5809	regulator of rDNA transcription protein 15	1.00E-74	Non-synonymous Ile(ATC)>Thr(ACC)
TgSNP_77	A/C	137	F: AGAACTGAATACATTCTCACAAATG R: CGTTGTTGGAGGGGCTAC	55.5	0.550	0.450	A=0.325	0.3042	not significant homology		Unkonwn
TgSNP_78	C/T	202	F: ACTGATTTCATAATTATCCTGGC R: CATTATTCTTCAAGAAAATATTCCAT	55.3	0.800	0.508	C=0.450	0.0082	not significant homology		Unkonwn
TgSNP_79	C/T	417	F: GAAAACCTCTGTAAGTTAATCTCTGA R: GGATGTACATGCTTTTTTATTAAAA	54.3	0.150	0.512	T=0.475	0.0012	PREDICTED: elongation factor-like GTPase 1 [Crassostrea gigas]	2.00E-50	UTR
TgSNP_80	A/T	152	F: CACTGGATTTCCTTTTTCACA R: GTTGAATTGCATGATGGGAT	55.3	0.100	0.097	T=0.050	0.8694	not significant homology		Unkonwn
TgSNP_81	A/G	65	F: TTTTCACCTTCCACCACGA R: ACCCTCTAACTGGATAACTTCAAC	56.1	0.600	0.513	A=0.500	0.4354	PREDICTED: angio-associated migratory cell protein [Pygocentrus nattereri]	2.00E-41	Non-synonymous Ile(ATA)=Met(ATG)
TgSNP_82	A/T	113	F: ATAATTGTATAATAACTTGTCATAATGTAAC R: TCTTATGAAAGTACAGGAACTTGTATAT	53.6	0.500	0.385	T=0.250	0.1609	not significant homology		Unkonwn
TgSNP_83	A/G	290	F: CAACATTCACACGGATTCATAA R: AACAAAACAATAATTACATCAACACA	56.2	0.400	0.492	G=0.400	0.3885	not significant homology		Unkonwn
TgSNP_84	G/T	160	F: GCTCGAGACACTCTGGACATT R: ACAGCATGGATGACGAAGC	56.5	0.350	0.358	T=0.225	0.9195	hypothetical protein CGI_10023086 [Crassostrea gigas]	1.00E-14	Non-synonymous[1] Asp(GAC)>Glu(GAA)
TgSNP_85	G/T	272	F: TGCCATCTTGAACAAACATAAT R: ATACATAGTCCCCTACCGGTC	55.3	0.600	0.467	G=0.350	0.1875	not significant homology		Unkonwn

TgSNP_86	A/G	813	F: TCAGTTGATCTCATTATATACAAGGTA R: GCAGGTTTGCTTAGATCTTATAGTT	53.6	0.500	0.467	G=0.350	0.7418	not significant homology		Unkonwn
TgSNP_87	G/T	228	F: CGAGTATGGAACCATCACGT R: GCTTCTTTCATCAACTGTCACTG	55.4	0.211	0.341	T=0.211	0.0770	PREDICTED: uncharacterized protein LOC 105337677 [Crassostrea gigas]	1.00E-17	Non-synonymous[1] Asp(GAC)>Glu(GAA)
TgSNP_88	A/C	138	F: CCTCTAAGTGTGTGAATTGTGGA R: TTCTGGTTAATATCTGATAATGTTGA	57.1	0.250	0.512	C=0.475	0.0190	not significant homology		Unkonwn
TgSNP_89	A/T	355	F: TCAGGTTGATACTCTGAGTAACATT R: TCTTTTTCCTCCTCTTCATCC	55.1	0.400	0.328	T=0.200	0.3004	WD repeat-containing protein 78	1.00E-44	UTR
TgSNP_90	C/T	194	F: CTCAATTCACCAATAATAAAGCAG R: TTTATTGGTTCAACCTTGGTGT	56.4	0.579	0.514	C=0.500	0.5682	not significant homology		Unkonwn
TgSNP_91	C/T	232	F: ATGACACTGTAATTTCAGATTCCTA R: AAAATCCCACAAAACTTCTACTT	55.2	0.350	0.358	C=0.225	0.9195	not significant homology		Unkonwn
TgSNP_92	C/T	339	F: GTGCATCGTCCAGTTTAGTTTT R: TTAGAGAGTTCTTTATGATGAACACTT	56.6	0.450	0.512	C=0.475	0.5809	not significant homology		Unkonwn
TgSNP_93	A/C	973	F: CACATTAGCTTAGTTTGAGGTTGA R: TTTGAATTAGAATGACAAAAATGTT	56.6	0.700	0.513	A=0.500	0.0940	not significant homology		Unkonwn
TgSNP_94	A/G	227	F: AATGATTACATAATCTCTCCATAAATT R: ATCCATAGGTCAGAGGTCATTTA	52.8	0.850	0.512	A=0.475	0.0024	glutamine synthetase	9.00E-161	UTR[1]
TgSNP_95	A/G	353	F: ATGTTTGTTCCTGTGTTCTGC R: CATTGCTTGAATTAAGTGTGTAAAT	55	0.500	0.492	A=0.400	0.9427	not significant homology		Unkonwn
TgSNP_96	A/G	301	F: TCTGTGTGTGTGTTTGTGCATA R: AAATATCCAATATAACAATGACTTTG	55.4	0.300	0.328	G=0.200	0.6841	not significant homology		Unkonwn
TgSNP_97	A/G	237	F: TCATGCATTTTTGGAGTAAATAA R: ACTTTGAAATGTGATGGATCTTAA	55.5	0.150	0.481	A=0.375	0.0015	not significant homology		Unkonwn
TgSNP_98	A/T	908	F: AGCAGCAAAATGATTATCAAGAT	56	0.200	0.185	T=0.100	0.6706	not significant homology		Unkonwn

			R: CATTAGCAATTACATGAACTATTTAGC								
TgSNP_99	A/T	399	F: TGTCCCATTTTCTACTACAAACTTT	57	0.368	0.309	T=0.184	0.3673	not significant homology		Unkonwn
			R: AAACATCTGTAATATCATAATGAGACAC								
TgSNP_100	C/T	176	F: TCATGGACAGACCCAATCAG	55.8	0.368	0.508	C=0.447	0.2186	not significant homology		Unkonwn
			R: CACTCTCTATGGTGCATACCACT								
TgSNP_101	C/T	1470	F: CTGTAACGTTCAAGGATAGAAATAA	55.6	0.650	0.512	C=0.475	0.2142	PREDICTED: uncharacterized protein LOC107350598 [Acropora digitifera]	3.00E-08	Synonymous
			R: GATTCGCCACACGACCAT								Gly(GGT)>(GGC)
TgSNP_102	A/G	242	F: TTTCCTGTTAAACCTTTTCTTCA	56.2	0.579	0.491	A=0.395	0.4197	not significant homology		Unkonwn
			R: TGATAAATCACGAAAGAGTTAAGAG								
TgSNP_103	A/G	58	F: GCAAGTAGAATATTTCAACTCGTC	55.5	0.667	0.508	G=0.444	0.1722	not significant homology		Unkonwn
			R: GTTTTTGTCTTGGAGATCTTCAC								
TgSNP_104	C/T	396	F: TTTCGGAAAATCTTTACATCTCA	56.7	0.500	0.467	C=0.350	0.7418	not significant homology		Unkonwn
			R: CTTACCAGTGTTCCACCTGATAC								
TgSNP_105	C/T	530	F: ATTCTAAAATGTTCTAAACCCAGAT	55.7	0.579	0.514	C=0.500	0.5682	not significant homology		Unkonwn
			R: TTTCAGTTGTTCAGTTGACATTTAT								
TgSNP_106	C/T	1015	F: TATCATTTTGACTCTCAAATCAATG	56.4	0.474	0.462	C=0.342	0.9117	not significant homology		Unkonwn
			R: GGCATAGCAATTATAATTAAGCAG								
TgSNP_107	C/T	110	F: TGCAAGAGTAATAAATAACACATCA	55	0.632	0.501	T=0.421	0.2411	not significant homology		Unkonwn
			R: CATTATGTATCACAACAACATGTGTA								
TgSNP_108	A/C	690	F: CCCAGACCACTCTACTTAAAATG	55.7	0.200	0.467	C=0.350	0.0084	not significant homology		Unkonwn
			R: GAACAACTGTAAATTTTTAGGCTG								

T_a, annealing temperature; H_o, observed heterozygosity; H_e, expected heterozygosity; MAF, minimum allele frequency; HWE, probability under assumption of Hardy-Weinberg equilibrium. UTR, untranslated region; [1], the contig containing SNP was the antisense strand, so the information of variant, SNP position were based on the antisense strand, and the information of the functional annotation, E-value, SNP location/effect were based on the sense strand