GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Development of novel polymorphic microsatellite markers for the blood clam *Tegillarca granosa* by pyrosequencing

Y.H. Dong, H.H. Yao, S.F. Shi, Y.B. Bao and Z.H. Lin

Zhejiang Key Laboratory of Aquatic Germplasm Resources,
College of Biological and Environmental Sciences,
Zhejiang Wanli University, Ningbo, China

Corresponding authors: Z.H. Lin / H.H. Yao
E-mail: zhihua9988@126.com / yaohanhan1020@126.com

Genet. Mol. Res. 14 (3): 8977-8987 (2015)
Received January 12, 2015
Accepted May 8, 2015
Published August 7, 2015
DOI http://dx.doi.org/10.4238/2015.August.7.6

ABSTRACT. Large amounts of expressed sequence tags (ESTs) generated using next-generation sequencing technologies provide a cost-effective and valuable genomic resource for the development of microsatellite markers. In this study, we isolated 115 novel polymorphic microsatellite markers for the blood clam *Tegillarca granosa* from ESTs in 454 sequencing data. All the loci were characterized in 30 individual clams from a natural population in Xiangshan (Zhejiang Province, China). The number of alleles per locus varied from 2 to 10, with an average of 3.78. The observed and expected heterozygosities ranged from 0 to 1 and from 0.040 to 0.799, respectively. The polymorphic information content (PIC) ranged from 0.038 to 0.825, and 29 highly polymorphic loci (PIC $\geq$ 0.5) and 42 moderately polymorphic loci (0.25 < PIC < 0.5) were identified. Thirty-eight of the 115 loci deviated significantly from the Hardy-Weinberg equilibrium ($P < 0.01$) after a Bonferroni correction. A BlastX search revealed that 46 (40%) of the polymorphic loci identified were from transcript regions of known genes. The microsatellite markers developed in the present study

will greatly enrich the microsatellite resources of *T. granosa*, and are available for further population genetic analysis, genetic trait mapping, and molecular-assisted selection.

Key words: *Tegillarca granosa*; Microsatellite; Polymorphic marker; Pyrosequencing

INTRODUCTION

The blood clam *Tegillarca granosa* L. is found in the Indo-West Pacific, from east Africa to Polynesia, north to Japan, and south to northern and eastern Australia. Its delicious taste and high nutritional value have made it a popular seafood, and it has been extensively farmed along the coasts of southern China and several southeast Asian countries. Although it is a commercially important maricultured species, there have been few genetic studies of *T. granosa* because of a paucity of effective molecular markers.

Microsatellite markers, also known as single-sequence repeats (SSRs), are one of the most powerful genetic tools available, and have been widely applied in diversity analysis, genetic mapping, association studies, and marker-assisted selection (MAS) (Duran et al., 2009). Although over 100 SSRs have been characterized in *T. granosa* to date (Gu et al., 2008; Liu et al., 2012; Dong et al., 2012, 2013; Zhou et al., 2013), the development of multi-allelic SSRs is currently inadequate for studies of genetic connectivity and genetic mapping.

Large *T. granosa* sequence data sets derived from large-scale expressed sequence tag (EST) discoveries have recently enabled the mining of SSRs (Dong et al., 2012). These EST-SSR markers are inexpensive, labor-saving, and are frequently associated with annotated genes. In the present study, we developed 115 validated polymorphic EST-SSR markers in *T. granosa* using deep transcriptome sequencing, which will be useful for future studies of this species in terms of its conservation genetics, evolutionary studies, and molecular breeding.

MATERIAL AND METHODS

Sample collection and DNA extraction

Thirty *T. granosa* individuals were collected from Xiangshan, Zhejiang Province, China (29.21°N, 121.46°E). The adductor muscles of all the samples were dissected and immediately preserved in alcohol at -20°C.

Total genomic DNA was extracted from the adductor muscles by standard proteinase K digestion, phenol-chloroform extraction, and ethanol precipitation. The quality and concentration of DNA were examined using 1.2% agarose gel electrophoresis and the OD_{260}/OD_{280} ratio, using a NanoVue™ UV/Visible spectrophotometer (GE Healthcare Ltd., UK). The DNA templates were adjusted to 100 ng/μL with ultrapure water.

Primer design, amplification, and genotyping

Microsatellite markers for *T. granosa* were characterized by screening the partial isotigs (ESTs) of our transcriptomic database (SRA052081) by 454 pyrosequencing. Putative EST-SSR markers were screened using the SSRHUNTER program (Li and Wan, 2005), which

was set for detection of di-, tri-, tetra-, penta-, and hexa-nucleotide motifs, with a minimum of 6, 4, 3, 3, and 3 repeats, respectively. Primers flanking the microsatellites were designed by the Primer Premier 5.0 software (Premier Biosoft International, USA), and synthesized by Sangon Biotech Co. Ltd (Shanghai, China).

Polymerase chain reaction (PCR) amplifications were performed in a 20-μL reaction mixture containing 0.5 U rTaq DNA polymerase (Takara, Japan), 1X PCR buffer, 0.2 mM dNTP mix, 2.0 mM $MgCl_2$, 0.25 μM of each forward and reverse primer, and approximately 100 ng template DNA. PCRs were conducted in a thermal cycler (Mastercycler® pro S, Eppendorf, Germany) under the following conditions: initial denaturation for 5 min at 94°C, followed by 35 cycles of denaturation for 45 s at 94°C, the optimized annealing temperature (Table 1) for 45 s, and extension at 72°C for 45 s, with a final extension at 72°C for 7 min. The PCR products were separated by electrophoresis on 8% nondenaturing polyacrylamide gel at 180 V for 3 to 4 h, depending on the size of the amplification product, and visualized by ethidium bromide staining under UV light. A 20-bp DNA ladder (MBI Fermentas, USA) was used as a reference marker to determine allele size. Genotypes were determined as approximate allele sizes.

Characterization of polymorphic EST-SSRs and sequence annotation

After optimization of the annealing temperature, primer pairs that produced clear bands of the expected size were selected for polymorphism analysis. Polymorphisms of each SSR locus were evaluated using 30 *T. granosa* individuals. All polymorphic SSR-containing ESTs were searched against the NCBI non-redundant (Nr) protein database and the Swiss-Prot database using BLASTx, with an E-value threshold of 1e-6. Gene names were assigned to each assembled sequence based on the best BLAST hit (the highest score).

Statistical analysis

The number of alleles, the observed (H_O) and expected heterozygosities (H_E), and the polymorphic information content (PIC) were calculated using CERVUS 3.0 (Kalinowski et al., 2007). Deviations from the Hardy-Weinberg equilibrium (HWE) for each locus were estimated using the Markov chain method by the online version of GENEPOP (http://wbiomed.curtin.edu.au/genepop/) (Rousset, 2008). All statistical significance levels were adjusted for multiple tests using sequential Bonferroni corrections (Rice, 1989).

RESULTS AND DISCUSSION

From a screen of 8805 ESTs that examined the length and guanine-cytosine content of the flanking regions, 246 primer pairs were successfully designed and used for validation of the amplification. After optimization of the experimental conditions, 157 primer pairs (63.8%) showed clear and consistent amplification patterns. Among these EST-SSRs, 115 were found to be polymorphic in the 30 clam individuals tested, resulting in a polymorphic percentage of 73.3%, which is similar to that reported by Dong et al. (2012) (76.5%), but is significantly higher than that found by Zhou et al. (2013) (46.8%).

Analysis of the 115 SSR-motif types showed that the most abundant repeat motifs were tri-nucleotides (51, 44.35%), followed by tetra-nucleotides (30, 26.09%), and di-nu-

cleotides (19, 16.52%) (Figure 1). This is consistent with the results of Zhou et al. (2013), but not with those of Liu et al. (2012). As is the case in other bivalves, penta-nucleotides (8, 6.96%) and hexa-nucleotides (7, 6.08%) were the least abundant motifs. Among the SSR motifs, TA (29.41%) and AT (29.41%) accounted for more than half of the di-nucleotide repeats, while CAA (9.80%) and TGT (7.84%) accounted for most of the tri-nucleotide repeats. AAAT (13.33%) and TTAA (10.00%) were the two most abundant tetra-nucleotide motifs.

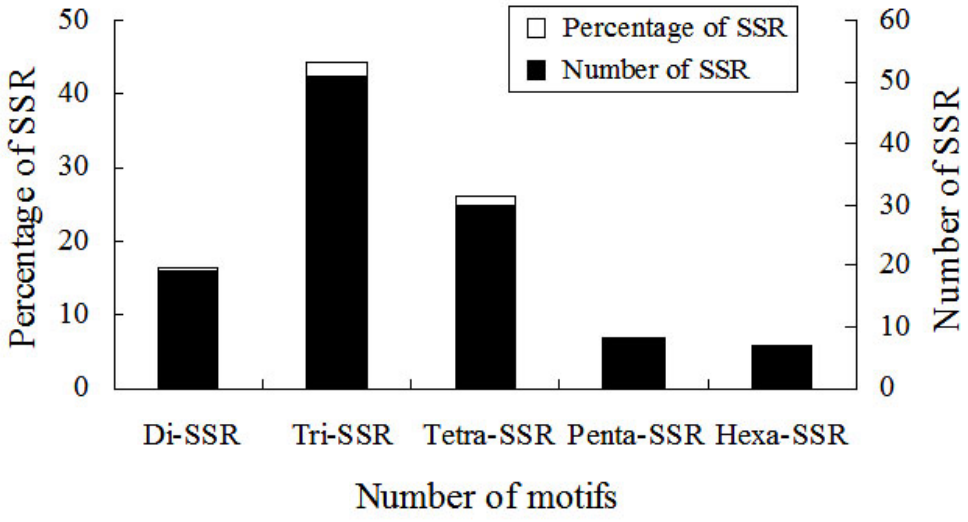

Figure 1. Distribution of 115 simple-sequence repeat (SSR)-motif types in *Tegillarca granosa*.

The number of alleles observed for the 115 loci ranged from 2 to 10, with an average of 3.78 (Table 1), which is similar to previous results obtained from EST-SSRs (Dong et al., 2012; Liu et al., 2012). The H_O ranged from 0 to 1, while the H_E varied from 0.040 to 0.799. The average H_O (0.3885) across all loci was lower than the H_E (0.4338).

The mean PIC was 0.3834, and ranged from 0.038 to 0.825. There were 29 highly polymorphic loci (PIC ≥ 0.5) and 42 moderately polymorphic loci (0.25 < PIC < 0.5) that were suitable for population genetic analysis. Thirty-eight of the 115 loci deviated significantly from the HWE ($P < 0.01$) after a Bonferroni correction (Table 1). This may be attributable to one or more factors, such as insufficient sample size, bottleneck effects, or even the presence of null alleles. Additionally, of the 115 polymorphic loci, 46 (40%) were identified from transcript regions of known genes that serve as type I markers, and can be easily linked to a phenotypic trait of interest, making them valuable for functional diversity studies (Varshney et al., 2005). Furthermore, the presence of SSRs in the 5'-untranslated region (UTR) can affect transcription or translation, while SSRs in the 3'-UTR can affect splicing (Li et al., 2004; Lawson and Zhang, 2008).

Overall, the transcriptome data of *T. granosa* using pyrosequencing contained a large amount of accurate sequence information that was suitable for the rapid and large-scale discovery of EST-SSRs. A total of 115 novel polymorphic SSRs for *T. granosa* were identified in the present study, which will greatly enrich the SSR resources of this species and provide a valuable tool for further population genetic analyses, the construction of linkage maps, and MAS.

Table 1. Characteristics of 115 polymorphic expressed sequence tag-single-sequence repeats (EST-SSRs) for *Tegillarca granosa*.

Locus	Primer sequence (5'-3')	Repeat motif	Ta (°C)	Size (bp)	N_A	H_O/ H_E	PIC	P-HWE	Functional annotations [a]/*E*-value	Accession No.
Tg11970	F: CATTTTTGTTCTGCCTCACT R: ATCCATCTCTAGGGACATAGTC	$(ATAC)_3$	61.1	111-115	2	0.040/0.040	0.038	1.0000	-	KF171182
Tg5455	F: TTTGCTGTTTTCCGACTCA R: GGGAGAAGACACGAAGGAG	$(ATTTT)_3$	63.1	190-198	2	1.000/0.510	0.375	0.0000*	-	KF171185
Tg5804	F: TTTCTCTATGCTCACTTTCCG R: TGACCTACAACGATACGACAAT	$(TAATAT)_3$	61.1	160-183	4	0.840/0.669	0.590	0.0001*	Superoxide dismutase [Cu-Zn]/4e-14	KF171184
Tg6987	F: TACTCTACGATTCCTCAAGCAA R: ATGACACACTTCTCTTCCTTCTTC	$(TGT)_5$	61.1	117-126	4	0.200/0.255	0.234	0.4051	-	KF171186
Tg6767	F: CGGGGTCATTTTGTATCGT R: ATGCTGGAGCGTCAGATTT	$(TTTAG)_4$	61.1	235-248	4	0.680/0.620	0.537	0.0000*	-	KF171183
Tg5938	F: GTAGGGTTTTTGCCCAGGTATTG R: CAACGGGAGAGCAGGAGATAGAA	$(TTG)_5$	63.1	185-190	2	0.040/0.040	0.038	1.0000	-	KF171187
Tg5972	F: TGACAGTTTGAAGACAGCAC R: CAGAGTTGGTTTTACTGTGG	$(AAT)_5$	63.1	210-228	5	0.480/0.407	0.372	1.0000	-	KF171188
Tg5996	F: TTACATTCGGTGGCGTTCGG R: ATCTTGTCGGTCTGTTTCGTTGGA	$(ACAACG)_3$	63.1	205-210	3	0.160/0.153	0.143	1.0000	Macrophage mannose receptor 1/4e-07	KF171189
Tg11521	F: GGTAGTCCAATAAAGTAGACCTGT R: CCCTGTAACTATGCGTGTGTA	$(AATA)_4$	64.5	175-188	2	0.680/0.458	0.348	0.0214	-	KF171190
Tg11456	F: TGTGATACCATCTCAGGAAGCGAA R: GATAGCCGTAACAGACGACAGAGC	$(CCA)_4$	64.5	195-210	4	0.480/0.699	0.630	0.0000*	-	KF171191
Tg11571	F: GTCTGAGAACTGACCAGGATTTA R: CATACAACTTACCCTTACGCAA	$(AAAG)_3$	64.5	180-203	6	0.200/0.443	0.414	0.0000*	-	KF171192
Tg11472	F: AGGATAATACTACAATCGGGTC R: TAGATTCGGTTACCTCAGATGT	$(CAA)_4$	64.5	120-133	4	0.280/0.349	0.320	0.2334	-	KF171193
Tg11614	F: CATCAGTGGAAGTCTACA R: ACTCGGAATACTTGAACT	$(TAAA)_3$	56.1	105-114	3	0.167/0.358	0.322	0.0002*	-	KF171194
Tg12255	F: ATCGGGGCTCCCATCATC R: AGAGAACGAGAAAACTTACCACTG	$(TAT)_4$	56.1	110-130	2	0.480/0.372	0.298	0.2775	Dolichol-phosphate mannosyltransferase/e-115	KF171195
Tg13782	F: CGGACAATCTCAAAAAACAG R: CCTTTGTTTGACCCTCTACTAC	$(TGA)_4$	61.1	165-170	2	0.000/0.490	0.365	0.0000*	-	KF171196
Tg6446	F: TTTAGGCGGTTCTACAACTACTTC R: GACATTTACATGGCTCCAGGAC	$(CTT)_4$	45.3	128-142	3	0.440/0.375	0.337	1.0000	Bromodomain adjacent to zinc finger domain protein 1A/3e-60	KF171197
Tg11881	F: TGCCACTACCACTACTGCTTCCAC R: ATCCCCAAAACAAGAACAGAAACG	$(CTT)_4$	45.3	242-258	4	0.880/0.647	0.570	0.0015	Cyclin-dependent kinase 11/8e-13	KF171198
Tg11241	F: GACCTGGACATTTTTCACACA R: TTCATCCTCCCAATCAGCA	$(TTTCA)_3$	45.3	240-255	4	0.917/0.675	0.604	0.0012	Programmed cell death protein 5/1e-24	KF171199
Tg12230	F: ATTCCATTCAAACAGCACT R: TAAGGCACTGATTGTAACG	$(ATG)_4$	61.1	200-220	5	0.560/0.636	0.570	0.0409	-	KF171200
Tg11282	F: TTGTCAAGATTGCTATGTGG R: TGGTTGGTGTGTTAGAGAAG	$(ATCA)_3$	53.4	220-243	3	0.040/0.153	0.143	0.0015	-	KF171201

Continued on next page

Table 1. Continued.

Locus	Primer sequence (5'-3')	Repeat motif	Ta (°C)	Size (bp)	N_A	H_O/ H_E	PIC	P-HWE	Functional annotations [a]/*E*-value	Accession No.
Tg11914	F: AGATGAGGTTGTCTTGGGAA R: CCTTTCAGGTGCTGGATTAT	$(AG)_6$	53.4	190-208	2	1.000/0.510	0.375	0.0000*	Protein CutA homolog/8e-35	KF171202
Tg12185	F: TCGGGAGCGACTAAGAAGAAGAT R: TGAAGTCAAGGGTCAAATGGAGAT	$(TTAA)_3$	64.5	238-263	4	0.440/0.457	0.413	0.5318	Cyclophilin F/3e-14	KF171203
Tg11340	F: TTGAATGCCATACAAGACC R: CGGATGTTAGAGTTAGCAAA	$(TAC)_5$	58.7	140-150	4	0.040/0.456	0.411	0.0000*	-	KF171204
Tg11476	F: CGGTTGATGGTGGATTTTT R: ACGGAGAACAATGGAAACG	$(TA)_7$	58.7	195-205	3	0.160/0.417	0.361	0.0001*	Dicarbonyl/L-xylulose reductase/2e-49	KF171205
Tg12190	F: TCATGGAACATGGTGTTGTA R: GCGAGATTAGTCAGGAAGAG	$(TGT)_5$	58.7	245-287	5	0.840/0.798	0.748	0.0004*	VEGF toxin/6e-19	KF171206
Tg12253	F: GGAATAGAAGCCCTGTTA R: CACAAGGCAGTGAATGAG	$(AAAT)_3$	58.7	248-278	3	0.040/0.153	0.143	0.0012	-	KF171207
Tg12204	F: GAATAAAAAAACAACGGGC R: ACAAGGTCAGATTACAGTTCGT	$(GT)_{11}(TG)_6$	58.7	120-162	7	0.783/0.817	0.771	0.0262	-	KF171208
Tg11971	F: ACGTTTCTGGCATTGTAGAA R: CATTGTTATGGAAGGGGG	$(TACA)_3$	58.7	118-132	3	0.600/0.530	0.444	0.0055	-	KF171209
Tg6284	F: CTTGCTCCTGTAGCCTCTG R: TACACACGAAAACGCCAA	$(TTC)_5$	45.3	185-223	5	0.667/0.565	0.508	0.3923	-	KF171210
Tg11363	F: CGAAACTGCAACAGAAGGA R: ATCTTCATCGGCAACCTGT	$(AAC)_5$	61.1	98-110	2	0.120/0.115	0.106	1.0000	Coiled-coil domain-containing protein/4e-55	KF171211
Tg11710	F: ATCCCACTTTCCTCTTCCTCAATA R: TACAACTTCTCCAACGACGACAAT	$(TGT)_5$	64.5	208-215	3	0.600/0.571	0.486	0.0013	-	KF171212
Tg11720	F: ATGACGAGAACCACAAGACG R: CAGAAATTCGACTGTAGGATG	$(GAATT)_3$	58.7	110-125	3	0.227/0.212	0.197	1.0000	-	KF171213
Tg11681	F: TCCTTTGTAACTTTCTCTCCA R: GGAAGAAGTTGCTTCAGAGA	$(TTTTC)_3$	58.7	140-152	2	0.125/0.120	0.110	1.0000	-	KF171214
Tg11685-1	F: TCAGCGGTTTGTCCTTTA R: GATTCTGGCATCTTGGTG	$(TTCC)_3$	48.5	100-119	3	0.200/0.187	0.172	1.0000	MGC107884 protein/1e-20	KF171215
Tg11700	F: GACAGTGGCAGTGACAAATG R: CTTCCTCTTCAGCAGACAAATA	$(TGA)_5$	63.1	177-190	3	0.040/0.381	0.317	0.0000*	TPR repeat protein 1/5e-34	KF171216
Tg13652-1	F: GTTTGTCAGTTTCGGCTTACC R: TGGAGGCAGTTTTTCCTTAGA	$(AAAT)_3$	64.5	170-214	7	0.480/0.802	0.754	0.0000*	-	KF171217
Tg6419	F: TCAACAGCAACAAAGACTTCAGA R: GCTTGATTTCTATTTTGTCCCTG	$(GGTCAA)_3$	58.7	122-138	2	0.560/0.490	0.365	0.6754	Yes-associated protein/1e-34	KF171218
Tg6144-1	F: AGTGATGGTGATTTGGAAT R: CTGCTTGTTTTGTTGCTT	$(GAT)_7$	61.1	164-199	7	0.520/0.589	0.548	0.0210	-	KF171219
Tg6762	F: TACAAGAAATCACAACTCGC R: TGGTGGTCAGGTTAAAATCT	$(TTAT)_4$	61.1	260-290	3	0.040/0.079	0.076	0.0187	Ubiquitin carboxyl-terminal hydrolase 34/2e-09	KF171220
Tg6144-2	F: TGATGAGGACAACGATGA R: CTGCTTGTTTTGTTGCTT	$(GAT)_5$	58.7	215-259	8	0.440/0.709	0.658	0.0000*	-	KF171219

Continued on next page

Table 1. Continued.

Locus	Primer sequence (5'-3')	Repeat motif	Ta (°C)	Size (bp)	N_A	H_O/ H_E	PIC	P-HWE	Functional annotations [a]/*E*-value	Accession No.
Tg5882	F: AACATCAAGGTCTGGTGC R: CCGCAGTGTCATCAGAAT	$(GAA)_4$	53.4	155-174	3	0.640/0.476	0.400	0.1423	Cat eye syndrome critical region protein 2/9e-50	KF171221
Tg6080	F: CCATCATTAGAATAGGTAACAGC R: GTTTTGACTCCTATGCTGACA	$(ATGTA)_3$	58.7	200-221	4	0.125/0.568	0.459	0.0000*	Acetylcholine receptor subunit beta/1e-37	KF171222
Tg6277	F: CACCAGTACAGTTTGTTATTTACG R: GAAGAAACCCCAACCGAC	$(GT)_{10}$	61.1	182-218	6	1.000/0.789	0.737	0.0000*	V-ATPase subunit E/5e-73	KF171223
Tg5736	F: CTATCAGAAAAGCCAGAATG R: GTCACCTGGACTTGGTATTAG	$(AAT)_4$	61.1	265-282	3	0.240/0.389	0.333	0.0079	Na(+)/K(+) ATPase alpha subunit/0	KF171224
Tg11685-2	F: GCGGATTTACAGTAATAGGGA R: ATTTGTGATGGATTAGTGACG	$(TTTG)_3$	58.7	114-138	5	0.720/0.582	0.519	0.0243	Phenylethanolamine N-methyltransferase/7e-20	KF171215
Tg13667	F: TCTTTGTCAGGTATTGGAGT R: GCAACTGTAGCAGCATCA	$(AT)_8$	48.5	158-198	4	1.000/0.618	0.525	0.0000*	-	KF171225
Tg11426	F: TAACATAATGCACTGGCTTG R: ATAATGTGATGTAGCAACGC	$(TTTA)_4$	48.5	218-227	2	0.320/0.509	0.375	0.1035	-	KF171226
Tg13681	F: CCAGGAAGTAGCACCACC R: CTTTAGTTGTTGCTGCTTTCT	$(CATCAG)_3$	61.1	150-158	2	0.120/0.115	0.106	1.0000	-	KF171227
Tg12596	F: TTCTGGTGGGAAAAACGACAA R: TTGCCGTCACCTATCCATTCT	$(AAAT)_3$	64.5	198-205	2	0.160/0.150	0.136	1.0000	Glycosyl hydrolase family protein/3e-13	KF171228
Tg11821-1	F: GTTGTGATCTTGGTCAAACTTT R: GTAGAAAAAACAAATGGTGGAC	$(TAT)_4$	61.1	148-178	5	0.440/0.656	0.585	0.0043	-	KF171229
Tg13405	F: CTGAGTTACAGAAGTTCAAACG R: AGTTTGCCCATCATCCAG	$(AAC)_4$	61.1	90-108	4	0.160/0.225	0.213	0.0724	-	KF171230
Tg11940	F: TGTGTCATTCTCTACTGCTG R: CTTCTTCCTCTTCTACCTGT	$(GAT)_4$	56.1	178-221	4	0.217/0.277	0.257	0.0687	-	KF171231
Tg12130-1	F: CAACATGCTGATGTGATGC R: CACTTTCAACTGTTTTCTCCA	$(TGA)_4(GAT)_5$	48.5	168-178	4	0.400/0.444	0.387	0.0016	FK506-binding protein 39/4e-09	KF171232
Tg13652-2	F: ATGTTTCACGGAAGATACCTGT R: GGTAAGCCGAAACTGACAAA	$(AAT)_5$	64.5	198-222	3	0.560/0.749	0.690	0.0004*	-	KF171217
Tg13652-3	F: GTATTTCCTCTGGGGTTTT R: CGTCTTTGTAGCGACATTTA	$(AATA)_3$	58.7	252-285	5	0.800/0.833	0.792	0.5207	-	KF171217
Tg7906	F: TGTGTTAGTGCTATCGTTC R: GTGTGCATCAACTACATATC	$(AG)_7$	61.1	161-184	5	0.680/0.618	0.553	0.0042	60S ribosomal protein L4/4e-11	KF171233
Tg12178	F: ACAGAGGAAGGTGAACAAACAA R: CACAACAGGTTCAGCAGGC	$(CAA)_5$	45.3	142-158	5	0.480/0.629	0.582	0.0000*	E3 ubiquitin-protein ligase Bre1/2e-55	KF171234
Tg13073	F: ATGCGTTGTAACTTTTGGAG R: AACAGTTGCCACCGATACA	$(ATG)_4$	64.5	214-223	4	0.400/0.562	0.498	0.0000*	-	KF171235
Tg13107	F: GTTTGTGGTACTAAATGGACTAC R: ATTTGCTACCGACATTGAA	$(TATC)_3$	61.1	132-145	3	0.040/0.079	0.076	0.0219	-	KF171236
Tg13334	F: CTATCCTCTTTTTTCCTACC R: AAGCAAAAACAGGATTGA	$(TA)_6$	56.1	133-155	4	1.000/0.706	0.635	0.0000*	-	KF171237

Continued on next page

Table 1. Continued.

Locus	Primer sequence (5'-3')	Repeat motif	Ta (°C)	Size (bp)	N_A	H_O/ H_E	PIC	P-HWE	Functional annotations [a]/*E*-value	Accession No.
Tg12173	F: TGTTTTCTTCTTCTCCACG R: GAGTTGCTTACCCTTTCGG	$(TAA)_4$	58.7	197-208	3	0.040/0.187	0.172	0.0011	-	KF171238
Tg13253	F: ACACTTCTCAATCAGCATAG R: GAAGTAGCCAATACCATCA	$(ATC)_4$	53.4	198-214	4	0.320/0.447	0.393	0.0414	Protein BCCIP homolog/2e-33	KF171239
Tg13145	F: TGTTACTTTATCGTTTGGC R: CAGTGACAAATCAGCAGAA	$(ATTTTT)_3$	56.1	108-114	2	0.360/0.301	0.252	0.5578	-	KF171240
Tg12849	F: CCAAGTTCAATGTATTCACC R: TTGTTTACTGTCATTCCTGC	$(TCA)_4$	61.1	218-226	3	0.320/0.287	0.262	1.0000	-	KF171241
Tg12973	F: ATCCACCTCACCACCTAAAC R: CAAATGAACCTCAGCCAATA	$(TCAGAA)_3$	63.1	182-192	2	0.680/0.458	0.348	0.0205	-	KF171242
Tg13411	F: ATCCGCTGATAAAGATGAAGAGGT R: TGTTGAAGGTTGCGTAGTAGTGGT	$(CAA)_4$	45.3	168-199	4	0.667/0.558	0.494	0.3370	-	KF171243
Tg13550	F: AACTACAGATTCAGGGACAA R: GTAGTTGTTGTCTTTGGTGTC	$(ATAC)_3$	61.1	174-185	3	0.040/0.220	0.199	0.0004*	-	KF171244
Tg13430	F: GAAACGAACAGGATGATTAG R: CAGACTTACTCTGAAAGACTTG	$(AT)_7$	61.1	120-142	4	0.200/0.320	0.295	0.0001*	-	KF171245
Tg11821-2	F: GAAAAATCACGCACAGCAAA R: CACTACAATGACGGCGATGA	$(AT)_7$	63.1	90-125	9	0.880/0.860	0.825	0.0000*	-	KF171229
Tg11830	F: TATGGCAACAAACACAGGAGA R: CGCTGGCTTGATTTTTTC	$(TCT)_5$	58.7	128-158	2	0.520/0.393	0.311	0.1428	-	KF171246
Tg12991	F: CATTCTGGCTTATGCCTTGA R: TGACTGGGACGTTAAACACC	$(TGAGT)_4$	56.1	133-145	4	0.240/0.396	0.348	0.0196	-	KF171247
Tg12130-2	F: TCATCTCCACCATCACTGCTG R: CAACATGCTGATGTGATGC	$(GAT)_{10}(GAT)_4$	48.5	120-148	7	0.440/0.711	0.651	0.0005	FK506-binding protein/4e-09	KF171232
Tg11854	F: TCGGGGTGTAAAAGGAAGTTG R: AGTGTGACAAATAATGGCAGACCT	$(TTTC)_3$	45.3	244-263	5	0.440/0.385	0.360	1.0000	-	KF171248
Tg13292	F: AGAGTTGTCAGGCGGAATG R: GCCGATTTACTGCTGAACTC	$(AGAA)_3$	46.6	194-202	2	0.167/0.156	0.141	1.0000	Armadillo repeat-containing protein 3/4e-47	KF171249
Tg13151	F: GAAGTTGAAAGATTGCTACAG R: CCACCTAAAAGGACTCATT	$(GT)_{10}$	58.7	228-253	7	0.400/0.787	0.742	0.0000*	-	KF171250
Tg13305	F: GTCTCAACCCAGTTCAAT R: AGCAGACACCTCTCTAAAA	$(TG)_{14}(AG)_7$	58.7	175-202	6	0.440/0.830	0.786	0.0004*	-	KF171251
Tg13614	F: TATCCACAAAGAGTTGAAGAAG R: GGTAAAGTGTAACCATTTGACA	$(ATT)_4$	45.3	235-257	4	0.667/0.588	0.517	0.0085	-	KF171252
Tg11714	F: GGTCACCAGAAAGAGAAGAT R: TGATACTTTCCTACGACGC	$(AT)_6$	58.7	232-248	4	0.880/0.712	0.543	0.0017	-	KF171253
Tg12108	F: TGAGGAAAATCTTGATGCTGACG R: ATCTTGTGTCTGCGGTGGGTTA	$(TGA)_5$	64.5	158-195	4	0.080/0.584	0.481	0.0000*	-	KF171254
Tg13617	F: TTGGTTTGGGATGGATACA R: GTGGTTATCTTGTAGGATTATTGT	$(TA)_6$	61.1	179-228	8	0.520/0.728	0.668	0.0000*	-	KF171255

Continued on next page

Table 1. Continued.

Locus	Primer sequence (5'-3')	Repeat motif	Ta (°C)	Size (bp)	N_A	H_O/ H_E	PIC	P-HWE	Functional annotations [a]/*E*-value	Accession No.
Tg13575	F: TGCCCCTATTCTTTTGCCT R: ATCCTCATCTTGACATCTTACAGC	$(AAGTC)_3$	64.5	258-284	8	0.360/0.650	0.604	0.0000*	Disco-interacting protein 2/4e-22	KF171256
Tg12309	F: GCAGTCACGTTTTTCTTCCCTT R: TTCACAACAGGTGGCTAATGGA	$(ATA)_4$	63.1	160-168	2	0.320/0.274	0.233	1.0000	Kv-beta-2/7e-59	KF171257
Tg13106	F: ATGTAAGCAAGACTACTGTATG R: GTGCTACAGTATTTGCTCTT	$(TTAA)_4$	61.1	198-218	2	0.200/0.184	0.164	1.0000	Atlastin-3/1e-41	KF171258
Tg13173	F: GCATCCGAAAAACTGGTA R: CATGAGGCTGATAATGGC	$(TTTA)_3$	58.7	214-230	3	0.200/0.489	0.424	0.0000*	-	KF171259
Tg12324	F: CCTTTTCCGACAACAATAA R: ATGTAATCTCTAGAGTGTACGC	$(CAAA)_3$	61.1	168-190	2	0.560/0.411	0.322	0.1319	-	KF171260
Tg12781	F: CCAAAAGCAAGAGGAAAT R: ACTAATCTACTGCCAAACTG	$(TTAA)_3$	58.7	185-210	3	0.080/0.079	0.076	1.0000	-	KF171261
Tg12479	F: CCAAGTTGATTTGATAGAC R: CAGTGATATTAAACACACC	$(TTA)_4$	53.4	92-108	3	0.080/0.222	0.205	0.0002*	-	KF171262
Tg12347-3	F: CACCACAAGCATCGCCAC R: CACTCGCTCATTCTTTCTTTTTCT	$(CAA)_4$	64.5	132-132	2	1.000/0.510	0.375	0.0000*	NimA-related protein kinase 7/7e-83	KF171263
Tg12657	F: TTAGAAGCAAGAAAACCTCAGC R: TTCTGTGGTAACCTGTGTGTATG	$(ACA)_4$	64.5	107-126	2	0.040/0.040	0.038	1.0000	-	KF171264
Tg12464	F: CAACCTGACAATCATACA R: CACCTACTAATCTGACTTT	$(AACA)_3$	58.7	248-261	3	0.200/0.189	0.176	1.0000	-	KF171265
Tg12328	F: ATGGCACTTATGTGAAAAA R: TGACACACACTAACGATAAAC	$(AT)_6$	58.7	96-108	4	0.080/0.118	0.113	0.0672	-	KF171266
Tg12347-1	F: AGTACCTGAACAAAGACCTGACATT R: GTGGCGATGCTTGTGGTG	$(ACA)_5$	64.5	178-202	4	0.040/0.190	0.181	0.0007	NimA-related protein kinase 7/7e-83	KF171263
Tg12412	F: TGGTCGTGGTTGTAGTCGGCTTT R: GAAAGAGACTGTCCCCCCTCCAAT	$(GTT)_4$	64.5	148-164	2	0.160/0.150	0.136	1.0000	-	KF171267
Tg12543	F: ATCCACGCAAATGTATTCTGT R: CCAGAGTAAACTTAGGTACAGCA	$(CCAT)_3$	53.4	178-198	2	0.720/0.470	0.355	0.0088	-	KF171268
Tg12806	F: CATGATGGAAAGTGGAGAAA R: AAAACCGATGCCCTGTGT	$(GATT)_3$	61.1	218-238	2	0.080/0.150	0.136	0.1224	-	KF171269
Tg11732	F: AAACAACTTCTGACCCATCTA R: GTAAAACCTGTAACTGTGCTAAC	$(AAAT)_3$	64.5	189-208	4	0.000/0.349	0.320	0.0000*	-	KF171270
Tg12347-2	F: TACCTGAACAAAGACCTGACATT R: TGCTTGTGGTGGCATCTG	$(AAC)_4$	64.5	168-201	3	0.040/0.528	0.403	0.0000*	NimA-related protein kinase 7/7e-83	KF171263
Tg12638	F: TTCACCTGCCTCATCTCC R: GGAAGACTTTTCAGAGTTTGTAG	$(AACACC)_3$	61.1	202-208	2	0.125/0.403	0.317	0.0017	Peptidyl-tRNA hydrolase 2/3e-30	KF171271
Tg12924	F: GCCAACACCAACTACATCTACAA R: ATCTGTCACCGTCATACTTCCA	$(ACA)_4$	45.3	199-211	3	0.200/0.187	0.172	1.0000	-	KF171272
Tg11742	F: ACTGACGCCTGTGAAGAAACTG R: AACATCAAGATTTTGCCAGCC	$(GTT)_4$	45.3	109-128	3	0.818/0.563	0.478	0.0177	Calcium homeostasis endoplasmic reticulum protein/4e-49	KF171273

Continued on next page

Table 1. Continued.

Locus	Primer sequence (5'-3')	Repeat motif	Ta (°C)	Size (bp)	N_A	H_O/ H_E	PIC	P-HWE	Functional annotations [a]/*E*-value	Accession No.
Tg12592	F: TAACTCTGCCTGTAGTCCGTATTT R: CTCCATCACCCTGAACATTTG	$(TGG)_7$	64.5	140-161	4	0.080/0.520	0.416	0.0000*	-	KF171274
Tg7970	F: GGAACAATAGACTCCACATAAACC R: TACCAAGACAACGCAAGCAC	$(TA)_7$	63.1	178-211	6	0.800/0.805	0.760	0.0000*	Collagen alpha-6(VI) chain/4e-47	KF171275
Tg11314	F: GCTGTTATGTGTGTTTCAAGA R: CTTGTCAGACTTTTGTCACTAA	$(TGT)_5$	63.1	152-158	2	0.000/0.327	0.269	0.0000*	MCM-binding protein/9e-59	KF171276
Tg12879	F: CTGCCACTTCCCAACCCAT R: GGACTTACTTCCTGCTCGCTG	$(CAA)_7$	64.5	189-202	3	0.528/0.403	0.403	0.0000*	Histone deacetylase 4/3e-17	KF171277
Tg11368	F: CCACTTGGATTTTCACACG R: GGTCTGTAACAAATGTCTGGG	$(ATAA)_3$	64.5	254-268	2	0.040/0.038	0.038	1.0000	-	KF171278
Tg3934	F: AATGCTGAGACCTTGTGAAC R: GTTTTGTATCCAAGACCTACAC	$(GT)_{17}$	59.5	238-278	10	0.600/0.841	0.801	0.0000*	Glutaredoxin-3/e-117	KF171279
Tg8139	F: AAGTTGGTACAAAGAAGTGCTG R: CCTGTATGGCTTAACATTATCAG	$(TG)_{21}$	61.1	227-258	7	0.600/0.804	0.759	0.0319	Syntaxin-5/3e-94	KF171280
Tg8561	F: GTTGGGCATTGTTTTTAGC R: CATTATCCGAAGATGTGTGG	$(AG)_7$	58.7	190-201	3	0.160/0.389	0.333	0.0009	Ribose-phosphate pyrophosphokinase 4/2e-71	KF171281
Tg8268-1	F: TTCAAACTCCCAGTCTTCAC R: AAAGGTTGCTATGGAAACTG	$(TCA)_8$	53.4	268-279	3	0.080/0.287	0.263	0.0000*	-	KF171282
Tg8772	F: AAGAACCTCCTTTGGCTAA R: CCTTTGTCTTCAGGTGTTGTA	$(AAG)_4$	50.8	214-242	4	0.480/0.408	0.374	1.0000	SH3 domain-containing kinase-binding protein 1/9e-42	JW036061
Tg8268-2	F: CACTGGTGATGTGACTAAT R: TGAAGATGTGGAAGGAGA	$(CTT)_4$	53.4	145-188	5	0.520/0.431	0.389	0.8761	-	KF171282
Tg8157	F: GATGAGAGAGACAAGATAGGAAG R: CTCTGTTAGAAGTTGTGCCATA	$(AT)_8$	61.1	264-288	5	0.680/0.767	0.708	0.0155	Di-N-acetylchitobiase /2e-59	JW036063
Tg7961	F: TATGGACACATCCCTGGGTTCTAA R: CCCTTATTTGAGTCACTTTCGCTG	$(AGT)_5$	48.5	192-228	4	0.160/0.154	0.147	1.0000	Transcriptional regulator ATRX/0	KF171283
Tg8197	F: CTCTGGTCCGTATGGAATGC R: GAAAAGGTCCTGTCCGAATC	$(TTTC)_3$	45.3	168-188	4	0.320/0.288	0.265	1.0000	Poly [ADP-ribose] polymerase 12/3e-21	KF171284
Tg8648	F: TTAGCACAGTTGGAATCAT R: CTCTATCAGAACTTTCCTCTTC	$(AGAA)_3$	58.7	110-122	3	0.040/0.079	0.076	0.0186	M phase phosphoprotein 10/9e-29	KF171285
Mean	-	-	-	-	3.78	0.3885/0.4338	0.3834	-	-	-

Ta = annealing temperature; N_A = number of alleles; H_O/H_E = observed heterozygosity/expected heterozygosity; PIC = polymorphic information content; P-HWE* = significant deviation from the Hardy-Weinberg equilibrium (HWE) after sequential Bonferroni corrections. [a]SSR-containing contigs were annotated using BlastX against Nr or the Swiss-Prot database, with an E-value cutoff of le-6.

ACKNOWLEDGMENTS

Research supported by the National High-Tech "863" Project of China (#2012AA10A400), the Zhejiang Major Program of Science and Technology (#2012C12907-4), the Modern Agro-Industry Technology Research System (#CARS-48), and Key Science Foundation of Zhejiang Province (#LZ12C19001).

REFERENCES

Dong YH, Yao HH, Lin ZH, Zhang LL, et al. (2012). Characterization of 62 polymorphic EST-SSR markers in the blood clam (*Tegillarca granosa*) and their cross-amplification in *Scapharca subcrenata*. *Conserv. Genet. Resour*. 4: 991-997.

Dong YH, Wu GX, Yao HH, Lin ZH, et al. (2013). Characterization of 34 polymorphic EST-SSR markers in *Tegillarca granosa* and their transferability in *Anadara craticulata*. *J. Fish. China* 37: 70-77.

Duran C, Appleby N, Edwards D and Batley J (2009). Molecular genetic markers: discovery, applications, data storage and visualisation. *Curr. Bioinformatics* 4: 16-27

Gu XY, Zeng QG, You ZJ and Lin ZH (2008). Isolation and characterization of six microsatellite primers of *Tegillarca granosa*. *Oceanol. Limnol. Sin.* 39: 661-664.

Kalinowski ST, Taper ML and Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16: 1099-1106.

Lawson MJ and Zhang L (2008). Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region. *Gene* 407: 54-62.

Li Q and Wan JM (2005). SSRHUNTER: Development of a local searching software for SSR sites. *Yichuan* 27: 808-810.

Li YC, Korol AB, Fahima T and Nevo E (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol*. 21: 991-1007.

Liu B, Teng SS, Shao YQ, Chai XL, et al. (2012). Isolation and characterization of 39 novel polymorphic EST-SSR loci for the blood clam, *Tegillarca granosa*. *Conserv. Genet. Resour*. 4: 375-378.

Rice WR (1989). Analyzing tables of statistical tests. *Evolution* 43: 223-225.

Rousset F (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103-106.

Varshney RK, Graner A and Sorrells ME (2005). Genetic microsatellite markers in plants: features and applications. *Trends Biotechnol*. 23: 48-55.

Zhou XL, Zhu JH, Dong YH, Lin ZH, et al. (2013). Development and comparative study of genomic-SSR and EST-SSR in *Tegillarca granosa*. *Oceanol. Limnol. Sin.* 44: 467-475.