J. Ocean Univ. China (Oceanic and Coastal Sea Research)
DOI 10.1007/s11802-017-3260-5
ISSN 1672-5182, 2017 16 (3): 509-516
http://www.ouc.edu.cn/xbywb/
E-mail:xbywb@ouc.edu.cn

Development and Validation of 89 Novel Expressed Sequence Tag-Derived Microsatellite Markers in Blood Clam, *Tegillarca granosa*

TENG Shuangshuang[1), 2), 3)], FANG Jun[1), 2), 3)], CAI Yilong[1), 2), 3)], CHAI Xueliang[1), 2), 3)], and XIAO Guoqiang[1), 2), 3), *]

1) *Zhejiang Mariculture Research Institute*, *Wenzhou* 325005, *P. R. China*
2) *Engineering Research Center for Marine Bivalves*, *Chinese Academy of Fishery Sciences*, *Wenzhou* 325005, *P. R. China*
3) *Zhejiang Key Laboratory of Exploitation and Preservation of Coastal Bio-Resource*, *Wenzhou* 325005, *P. R. China*

(Received July 13, 2016; revised October 8, 2016; accepted March 13, 2017)

Abstract Blood clam, *Tegillarca granosa*, is an important shellfish in Chinese mariculture industry. Investigative research in this species, such as genetic linkage mapping, requires a large panel of molecular markers. In present study, a total of 89 polymorphic microsatellite markers were developed in *T. granosa* using the sequence database of Life Sciences Technology 454 next generation sequencing technology. All 89 loci were characterized in 20 individual clams from a natural population inhabiting Yueqing Gulf, Zhejiang Province, China. The number of alleles per polymorphic locus varied between 2 and 15, while the observed heterozygosity, expected heterozygosity and polymorphic information content varied between 0.000 and 1.000, 0.102 and 0.921, and 0.048 and 0.886, respectively. Of the 89 loci identified, 32 loci deviated significantly from Hardy-Weinberg equilibrium following Bonferroni correction. Thirty nine markers, which were shown to be polymorphic in a full-sibling family, were tested in Mendelian segregations. As expected, 32 loci were co-dominantly segregated in a Mendelian fashion. These novel developed microsatellite markers represent useful research tools for investigation of population genetic structure and genetic diversity in this species.

Key words *Tegillarca granosa*; expressed sequence tag; microsatellite

1 Introduction

Blood clam (*Tegillarca granosa*) is a commercially important species of shellfish in Chinese mariculture industry and is extensively cultured from southern Korea to Malaysia, particularly along the southern Chinese coast (Broom, 1983; Wang *et al*., 2008). Between 2011 and 2014, the total aquaculture production of clams in China averaged 315000 tons per year. *T. granosa* is sold extensively along the coasts of southern China and several Southeast Asian countries. China has a long history of *T. granosa* aquaculture, and artificial breeding technology in *T. granosa* was solved in 1990s. The artificial breeding and culture in *T. granosa* develops rapidly and the marine environment becomes polluted over the last two decades. Consequently, the degradation of genetic resources of *T. granosa* populations has occurred, which can have detrimental consequences on *T. granosa* farming. In order to improve breeding efficiency, the genetic conservation of *T. granosa* germplasm resources has become more urgent for the sustainable management. Genetic studies, which offer great potential to detect associations between allelic forms of a gene and phenotypes, will improve the development of the *T. granosa* industry. Molecular marker assisted selection (MAS) technology, which is one of the breeding technologies, is a process that links molecular markers to targeted phenotypic traits to facilitate the screening and identification of individuals with beneficial phenotypic traits. One particularly useful approach for the use of MAS in breeding programs is to construct a genetic linkage map and subsequently map quantitative trait loci (QTL). Therefore, it is very important for the continued success of *T. granosa* culture to develop a large number of polymorphic molecular markers to construct high density genetic linkage maps.

The 454 high throughout sequencing technology has been proved to be cheap and efficient for detecting DNA sequences. It has been widely used in plant genome sequencing, genotyping analysis, transcriptome sequencing, and methylation research (Wicker *et al*., 2006; Macas *et al*., 2007; Wang *et al*., 2011; Bowman *et al*., 2012; Rahmann *et al*., 2013). The 454 transcriptome sequencing database has allowed the rapid identification of microsatellites from a variety of species including the copperhead snake

* Corresponding author. Tel: 0086-577-88210966
E-mail: xiaogq1978@163.com

(*Agkistrodon contortrix*) (Castoe *et al*., 2010), blue duck (*Hymenolaimus malacorbyncbos*) (Abdelkrim *et al*., 2009), and mungbean (*Vigna 510adiate*) (Sithichoke *et al*., 2009). Microsatellites or simple sequence repeats (SSRs), as one of the most popular genetic markers, have been extensively applied to genetic linkage mapping, germplasm resource genetic analysis, and parentage identification. Microsatellites have proved to be highly polymorphic, co- dominant, abundant in genomes and genotyping easy. To date, approximately 211 microsatellites have been developed for *T. granosa* (Liu *et al*., 2012; Dong *et al*., 2012, 2013; Gu *et al*., 2008; Shi *et al*., 2013; Zhou *et al*., 2013), containing 23 genomic-microsatellites and 188 EST-de- rived microsatellites. Collectively, all these microsatellites offer a robust resource for *T. granosa,* with which the germplasm resources of both wild and cultured stocks are estimated and compared. However, existing data are not sufficient for microsatellites-based mapping construction studies such as the identification and mapping of quantitative trait loci (QTL) and marker-assisted selection (Hubert and Hedgecock, 2004).

In the present study, we developed a set of EST-derived microsatellite markers for *T. granosa* from our existing Life Science Technology 454 sequencing database. We successfully identified 89 novel EST-derived microsatellite markers derived from a transcriptome dataset, which will allow the detection of genes responsible for desirable quantitative traits in a manner that is much more efficient than using genomic sequences (Muchero *et al*., 2011).

2 Materials and Methods

2.1 *T. granosa* and DNA Extraction

To screen polymorphic microsatellites, 20 natural *T. granosa* (shell length: 30 mm ± 1 mm) were collected from Yueqing Gulf, Zhejiang Province, China. To identify the inheritance patterns of polymorphic microsatellites, a full-sibling family was used to ascertain if alleles were inherited in a Mendelian fashion. This sibling family consisted of two parents originally produced in 2007, and 50 progeny (two years of age at the time of study). Genomic DNA was extracted from the adductor muscle each mollusk by standard proteinase K digestion, phenol-chloroform extraction, and ethanol precipitation as described by Li *et al*., (2002). The quality and concentration of the DNA extracted were examined by 1.5% agarose gel electrophoresis and determination of OD260/OD280 ratio using a Nanodrop 2000 UV/visible spectrophotometer. All DNA templates were adjusted to 100 ng μL^{-1} as working DNA templates with TE buffer.

2.2 Data Mining for Microsatellites Markers and Primer Designing

Putative microsatellite markers for *T. granosa* were screened from unigenes within our Life Sciences Technology 454 sequencing database. In total, 3100 unigenes were selected for screening purposes. Of these unigenes, 222 microsatellite-containing unigenes were identified using SSRHUNTER software (http://www.bio-soft.net/dna.html). Primers were then designed for these 222 microsatellite-containing unigenes each using Primer Premier 5.0 program (http://www.premierbiosoft.com/). Primers were designed according to the following criteria: 1) primers in lengths of 18 bases at least; 2) the amplicon size preferable around 200 base pairs; 3) the annealing temperature of upstream and downstream primers similar and between 50 and 65℃; and 4) GC content between 40% and 60%.

2.3 PCR Condition

PCR amplifications were performed in a 20 μL volume including 0.5 units r*Taq* DNA polymerase (Takara, Japan), 1× PCR buffer, 0.2 mmol L^{-1} dNTP (each) mix, 2.0 mmol $L^{-1}MgCl_2$, 15 μmol L^{-1} forward and reverse primers (each), and 100 ng template DNA. PCR was performed using the following conditions: predenaturing at 94℃ for 5 min, followed by 35 cycles of denaturing at 94℃ for 45 s, annealing at temperatures appropriate for primer pairs each (Table 1) for 45 s, and extending at 72℃ for 45 s, and a final extension at 72℃ for 5 min. Amplification products were separated on 6% denaturing polyacrylamide gels and visualized by silver staining as described by Liu *et al*. (2004). A 10 base pair DNA ladder (Takara, Japan) was used as a molecular weight reference in order to determine the size of PCR amplicon.

Table 1 Characteristics of the 89 novel microsatellite markers derived from *Tegillarca granosa*

Locus	Primer sequences (5'-3')	T_a (℃)	N_a	H_o	H_e	PIC	HWE	Genbank accession
Teg3902	F: CATATTTTCCAAGGACACAAA	50	8	0.842	0.861	0.818	0.002	KU293596
	R: TTTCAGAAAATACTTGACTACC							
Teg3912	F: TGTGAGTAACAAGGCCATAAT	48	2	0.316	0.444	0.339	0.192	KU293597
	R: CACGACAGGATATTTCACAAT							
Teg3930	F: AATCGCTAAAGTGATATGCAA	50	5	0.111	0.260	0.245	0.000	KU293598
	R: AGACTTTGTTGCTTTTCTTCC							
Teg3967	F: GTGTTACCATGGCAATATTTTA	50	11	0.526	0.909	0.874	0.000	KU293599
	R: CCTTGATAAAAACAAGTGAGC							
Teg3991	F: TGAATCCAAAACTTTCAGGTA	50	7	0.500	0.613	0.566	0.705	KU293600
	R: CTTCCGAAAACAGTCCTTATT							
Teg4020-1	F: ATGACACGCATGACATACATA	50	10	0.529	0.854	0.807	0.318	KU293601
	R: GCCTTGAAAATCCACATATAA							

(*to be continued*)

(*continued*)

Locus	Primer sequences (5'-3')	T_a (℃)	N_a	H_o	H_e	PIC	HWE	Genbank accession
Teg4020-2	F: GGCCCTTGTTTACAAACTACT	50	2	0.421	0.341	0.277	0.282	KU293601
	R: AAATTAAAAACACCACGGAAG							
Teg4086	F: CCATTTCCTATTTCACAGTTTT	50	6	0.450	0.582	0.534	0.625	KU293602
	R: GCAACCCTGAAGTTCTTTATT							
Teg4093-1	F: ACATATTTTACGGTTTCCACA	50	6	0.313	0.698	0.618	0.010	KU293603
	R: GTTGTACAGTTGCTTTGAAAAA							
Teg4093-2	F: AACATCCATTGCCATGTAATA	50	2	0.000	0.102	0.095	0.000	KU293603
	R: GTTAAAAACGTGCAAAAATG							
Teg4121	F: GCAACTACATCAATTCATCATC	50	3	0.842	0.569	0.478	0.025	KU293604
	R: GGCAAGGTCAAAGTAGACTG							
Teg4170-2	F: AAGGTTGTTGTTGTTGTTGAT	50	2	0.790	0.491	0.364	0.006	KU293605
	R: CAGAAGAAACCACTGTTATCG							
Teg4227	F: TAGGACCTTGATTTGTTTTTG	50	3	0.412	0.355	0.313	0.812	KU293606
	R: GACTGTTGTATTTATTGGCATTT							
Teg4294	F: AGAGTGGTCTCCCTTCATATC	53	9	0.895	0.774	0.722	0.002	KU293607
	R: CCACTATGGTTTAAGGGATTT							
Teg4303	F: TAGGGCATTATAAATGGTTCA	50	4	0.750	0.550	0.438	0.000	KU293608
	R: GGCTTAAAACACTCCAAAAAT							
Teg4356	F: TTTTACTTGGGGTAAACATTG	50	2	0.000	0.097	0.090	0.000	KU293609
	R: AACTTCAAAGAGACCATTTCC							
Teg4382	F: AGGACAGTATGCTGACAAAAT	53	6	0.071	0.675	0.613	0.000	KU293610
	R: GGTATTTCAGTTGGCAAAAAT							
Teg4458	F: CGAATCGAAAGTTTGTATAGG	45	4	0.474	0.460	0.399	0.000	KU293611
	R: AATATTTGGTCAAGCACACAC							
Teg4478	F: GAAGTTCAATGACAGCAAATC	48	5	1.000	0.747	0.691	0.004	KU293612
	R: TTAGCTCCTGTGTCTCAGATG							
Teg4485	F: CTTGATGAGAACAACTGATGC	45	4	0.667	0.668	0.583	0.945	KU293613
	R: GGTTTATTGTGTGGTTGGATA							
Teg4495	F: CAGGTACCATGAATCTTCAAA	50	2	0.368	0.371	0.296	0.972	KU293614
	R: CATTCTGGCTGAATCTATCAC							
Teg4516	F: GATCGTCTTTTTCATCTTCAA	50	3	0.105	0.152	0.142	0.000	KU293615
	R: AGAATTTCATCATGTATGGTTG							
Teg4540	F: AGATTGGAACAGTCGTTACAG	50	5	0.368	0.576	0.491	0.475	KU293616
	R: TAATGAAACCTCGTGCTAAAC							
Teg4542	F: TTTATCACAGCAGTTATGCAG	50	7	0.611	0.811	0.758	0.459	KU293617
	R: GCCACGGAAGAACAATATTAC							
Teg4547	F: CTTGTGGAAATTTTCTTGCTA	50	2	0.053	0.235	0.202	0.000	KU293618
	R: TCTGACAAATATTTCACGTCAT							
Teg4613	F: TTAACAATGACCATTTCATCC	50	4	0.737	0.630	0.561	0.300	KU293619
	R: TTGTTATACTGGCTTTTGCTC							
Teg4649	F: GAACTGGTGAACAAAAACAAA	55	3	0.800	0.528	0.424	0.041	KU293620
	R: TAATTGTCACGAACACACTGA							
Teg4654	F: TGGAATTGAGTGTCCTAAAAA	50	2	0.177	0.451	0.342	0.009	KU293621
	R: CAGAATTGTTTATGGGGATTA							
Teg4667	F: GGGAAATTACACGAAATTGTA	50	5	0.333	0.749	0.684	0.000	KU293622
	R: CTCATTAAGAAATGTAAAAGACG							
Teg4702	F: TTTCGAGGTATTTGTTTCTGA	50	6	0.684	0.724	0.656	0.682	KU293623
	R: TGTTTTCCGATGGTTATGTTA							
Teg4719	F: AGTGGTGACTACCATTAAACG	50	6	0.611	0.608	0.564	0.440	KU293624
	R: CATGAGAAAACTAAGGGGTTT							
Teg4730	F: GGGAACTCGTAGTTCTAAAGC	50	2	0.000	0.102	0.095	0.000	KU293625
	R: ACATTTAGACATTGGAAACCA							
Teg4783	F: CTGCCCAATTTAAGTTTTAGA	50	2	0.000	0.097	0.090	0.000	KU293626
	R: TCTACTCAAATGGTTGTGACC							
Teg4806	F: TTCCAGCATCAACAAAATAAC	50	4	0.222	0.462	0.418	0.000	KU293627
	R: AGCACATTAGGGGAAATTAGT							
Teg4849	F: TTTACCAACGAAGTCAAAGTG	50	3	0.790	0.605	0.510	0.029	KU293628
	R: CGAAATCTGTAATTGAAACTCA							
Teg4854	F: TGTGGCACATATTATTCTGCT	50	3	0.790	0.619	0.519	0.246	KU293629
	R: GACTCAATTTGACTGTTCTGG							

(*to be continued*)

(*continued*)

Locus	Primer sequences (5'-3')	T_a (℃)	N_a	H_o	H_e	PIC	HWE	Genbank accession
Teg4859	F: CGAATCTCCATCTTCCAAT R: AGACCCCAGAAGTTACAGAAC	50	2	0.000	0.102	0.095	0.000	KU293630
Teg4876	F: TAGGGACTTGCATCTTCTGTA R: GTCATGCTTTAAACTAACACACA	50	5	0.412	0.763	0.697	0.017	KU293631
Teg4893-1	F: TCCATAGCCTTCAGCATAATA R: TGGTCAATAAGGCTGATAAAA	50	4	0.450	0.481	0.428	0.973	KU293632
Teg4893-2	F: TTATCAGCCTTATTGACCAAA R: TCCACAATCAATTTCATGC	50	5	0.737	0.809	0.754	0.323	KU293632
Teg4953	F: TACACCTTGTCCCTACATCAA R: TTAGTCATTGTTTTAGTGAGGT	50	4	0.684	0.569	0.515	0.633	KU293633
Teg4996	F: ATTCCAATAAATGGCAAGC R: TAAATGACCCATGAAACAGAC	50	3	0.368	0.489	0.424	0.304	KU293634
Teg5082	F: CTAAAATGTTCAACCCAGTTT R: CCAGAATAGTTTGACTTCATCA	50	3	0.526	0.607	0.524	0.116	KU293635
Teg5137	F: GCAGTGAGAAGATGTTTCTGT R: TATTCATCTTCGTCCCAGATA	50	3	0.067	0.467	0.393	0.000	KU293636
Teg5150	F: TCTTTTGCAATCGTTTTTAG R: ACTGGTCATTGATGATTGTTA	50	3	0.500	0.586	0.505	0.037	KU293637
Teg5195	F: TGGTGCAATATCAAATAAACTG R: TCTGCAGGAAGAAAAACATAA	50	5	0.421	0.730	0.660	0.014	KU293638
Teg5200	F: CATCATGACCAATACAACACA R: CTGGTTGTTGATATAGCAGGT	50	3	1.000	0.578	0.469	0.000	KU293639
Teg5219	F: GATTGTCTCCCTTTGAAGTGT R: CTTCTCAAGCAAAATGTAACAA	50	12	0.947	0.868	0.829	0.431	KU293640
Teg5262-1	F: ATTGTAGTCCCATTGGTAGAT R: CACGAAAATAGTGGAAGTCAA	50	3	1.000	0.539	0.412	0.000	KU293641
Teg5262-2	F: CCTCCTTTTCATCATTACACC R: AAACAGCATATCAAGAAAACG	50	5	0.444	0.533	0.485	0.386	KU293641
Teg5286	F: TATGTAGCTGTTGCACAATGA R: GCAACAGGTGACACAAATTAC	50	5	0.529	0.754	0.692	0.046	KU293642
Teg5366	F: AATCCATGCAAACATTGTAAG R: CAAGTGAATTGTGTTCATCTG	50	3	1.000	0.640	0.553	0.000	KU293643
Teg5369	F: TGATTTACGCTAACAAATTGAG R: CCAAAACTCTACCTCTCTTTCA	50	5	1.000	0.627	0.534	0.032	KU293644
Teg5370	F: TGTCATCTTGTGAGACAAACTT R: GTGTGTGTGGGTGGGTAG	50	6	0.944	0.662	0.577	0.000	KU293645
Teg5380	F: ACTATTGCATTTCCTCATGG R: AAGCTTGTGTATGTGTGTGTG	50	8	1.000	0.727	0.661	0.157	KU293646
Teg5416	F: GCAATAACTAGCGATAGATGC R: GAAGGTGTGATGAAAGAAGAA	50	5	1.000	0.689	0.610	0.000	KU293647
Teg5440	F: GCCAACTACTTCAACAACAAC R: CCTTCTGTTGTCTGGGTTAAT	50	6	1.000	0.762	0.700	0.001	KU293648
Teg5468	F: ACGCGAGTTATGTTTTTATCA R: ACAGGTGTGTCGAGTTTAACA	50	9	1.000	0.821	0.772	0.045	KU293649
Teg5470	F: ACATGCATAATTTCACAGGAT R: TGTGTAACTTGTTTCATTGTCA	50	5	0.500	0.775	0.713	0.000	KU293650
Teg5478	F: ATAGGTTCATCTCACAAAGGA R: GACAGGTAAGCCTGCTCTTAT	50	4	0.842	0.552	0.441	0.000	KU293651
Teg5507	F: GTGTAAACTCCAACCAGT R: TTGAAGCAACAACAATCT	50	9	0.471	0.822	0.774	0.000	KU293652
Teg5572	F: CATTAACCTGACCTGCTA R: GGTGAACTGCTTCATTAT	50	4	0.947	0.588	0.479	0.000	KU293653
Teg5627	F: CCACTTTGGTCATTCTTC R: GTGCTTTACATGACAACATC	50	4	0.895	0.707	0.629	0.009	KU293654
Teg5660	F: TGATTGGGTCACAAAATA R: TGAAAGACATAGCGAAAA	50	4	0.105	0.371	0.342	0.000	KU293655
Teg5790	F: ACAACCAAAATAACAACATC R: CAAGGGGTGGAGGACTGC	50	4	1.000	0.607	0.506	0.005	KU293656
Teg5791	F: GTTTGCTGTGCTCAGGTG R: CCATTATTGGAGTTTACTTTCAT	50	5	0.684	0.589	0.482	0.947	KU293657

(*to be continued*)

(*continued*)

Locus	Primer sequences (5'-3')	T_a (℃)	N_a	H_o	H_e	PIC	HWE	Genbank accession
Teg5859	F: GGCCATATTTTATGTTCT R: GGTACCTGTTTGGTTTTA	45	2	0.333	0.286	0.239	0.444	KU293658
Teg5870	F: GAGCCAAGCAGTCCTAGT R: CTTCTTCCTCCTCCATCT	50	6	0.895	0.700	0.635	0.956	KU293659
Teg6017	F: GAGACAACAAACACCGAC R: TGTGAAAGCAGAAGCAAT	58	2	0.500	0.385	0.305	0.161	KU293660
Teg6027	F: CAATCCCAGAAAATTCGA R: TCTACCTAACAAAGCATAA	50	3	0.000	0.610	0.528	0.000	KU293661
Teg6162	F: AAGGGTGATTATACTATA R: AGGTACAAATTAACAAAG	50	6	0.833	0.770	0.708	0.056	KU293662
Teg6169	F: TTTTGTTCTTTTGTTAGT R: AGATGTCTTTTATGTAGG	45	3	0.833	0.560	0.445	0.000	KU293663
Teg6173	F: GTCAAACTGGTCAGAATA R: GCAAAAAACTAACCCTAA	50	8	1.000	0.820	0.768	0.000	KU293664
Teg6243	F: TTGTGACTTATTTCTCTC R: TGTAATGTGTTAATGCTT	50	4	0.263	0.371	0.342	0.006	KU293665
Teg6371	F: CTAACCATCAAAACAAAA R: AACTTTCACGAACAACTT	50	7	0.615	0.868	0.813	0.080	KU293666
Teg6573	F: AATGAAATGGGATAAAAT R: TACTGGGCTGGTGCTCTG	50	4	0.263	0.289	0.267	0.066	KU293667
Teg6645	F: GTGTAATACTTGGTCGGT R: AAACTTCAATATAGCTCA	50	5	0.188	0.434	0.400	0.000	KU293668
Teg6682	F: AAAGCACTGTGACAGAGA R: GCATGAAAATTGGATTAG	50	3	0.450	0.432	0.365	0.873	KU293669
Teg6727	F: AAATCGTGTAACAAATCT R: AACATAGCATAACATAAC	50	15	0.889	0.921	0.886	0.793	KU293670
Teg6790	F: TGAAAAGAATGAATGTTA R: TTGATTTGAAAAAAGTTA	50	3	0.050	0.145	0.136	0.000	KU293671
Teg6808	F: CACACACGATACAAGCAC R: AATAATAAGGAGGCACAG	50	11	1.000	0.917	0.870	0.036	KU293672
Teg6833	F: GATTTACATCTCCAAAGT R: GGCAATATCAAAGACCAT	50	6	0.533	0.743	0.677	0.099	KU293673
Teg6841	F: TATTTTCTGACTTTTATC R: AATGTAGTATTGTATGTG	50	2	0.071	0.071	0.066	1.000	KU293674
Teg6847	F: AAGTGTTTAGTGTGTGTT R: CAAGTACTGGTATTTAGC	50	4	0.350	0.317	0.294	0.993	KU293675
Teg6872	F: ATACTTATACATACCAAA R: TAGATTATCTAGCACTTA	45	9	0.857	0.897	0.850	0.011	KU293676
Teg6915	F: TCATGTTGATGTTACTTA R: CTACACTGGTTCCGTTAT	50	2	0.050	0.050	0.048	1.000	KU293677
Teg6978	F: AAATTGTCTTGTTCTTCG R: GGCACTTTATCATTACCA	42	11	1.000	0.897	0.858	0.246	KU293678
Teg6981	F: GTTTTAGAATAGGACCAA R: CCTGTAATATGATTACCG	50	6	0.550	0.640	0.592	0.397	KU293679
Teg7001	F: ATACCCAAGGGAACTAAT R: ATACCCATCAACGACACA	53	3	0.850	0.522	0.402	0.017	KU293680

Notes: T_a, annealing temperature; N_a, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; PIC, polymorphism information content; HWE, Hardy-Weinberg equilibrium. *, Significant deviation from Hardy-Weinberg equilibrium after Bonferroni correction for multiple tests (k = 89).

2.4 Statistical Analysis

The number of alleles (N_a), the observed (H_o) and expected heterozygosity (H_e), and tests for linkage disequilibrium and deviation from Hardy–Weinberg equilibrium (HWE) of these microsatellite markers identified were calculated using the GENEPOP program (http://genepop.curtin.edu.au/) and are displayed in Table 1. Significant levels were calculated per locus using Bonferroni method (Rice 1989). Polymorphism information content (PIC, Table 1) was estimated using the CERVUS 3.0 program (http://www.softpedia.com/get/Science-CAD/Cervus.shtml). To check the consistency of all the microsatellites to Mendelian segregation, the inheritance pattern of polymorphic markers was analyzed in a ful-sibling family with two parents and fifty offspring. Chi-square analysis was used to measure all observed progeny ratios against the expected Mendelian segregation ratios (1:1, 1:2:1 and 1:1:1:1) at each polymorphic microsatellite loci at the 0.05 probability level.

3 Results and Discussion

3.1 Isolation and Design of Primers for Target Microsatellites

We randomly selected 3100 unigenes from our existing Life Sciences Technology 454 sequencing database, from which a total of 222 microsatellites-containing unigenes were identified. The sizes of the microsatellites derived were not even. Dinucleotide repeats were the most abundant and accounted for 62.16% (138 loci). Trinucleotide, tetranucleotide, and compound nucleotide repeats were found at much lower frequencies (54 loci, 11 loci and 19 loci, respectively). Of the 222 microsatellites-containing unigenes that were identified, 153 primer pairs were designed successfully. There were several reasons for our inability to design a full complement of primers. Firstly, there were less than 50 nucleotides on either end of the simple repeat sequences, thus preventing the Primer Premier software from designing appropriate primers. Secondly, we experienced several cases in which there were hairpins, dimers, false priming, and cross dimerization between the forward primer and reverse primer, thus preventing such primers from being used for PCR amplification.

3.2 Characteristics of Microsatellites

We selected all 153 microsatellite loci for which we had successfully designed primers and used these for optimization experiments about whether the PCR products were as expected length and the anneal temperature of each primers was optimal. Our initial tests showed that 94 of these loci were successfully amplified at the expected length, while 59 loci were proved difficult to amplify. The unigenes obtained from transcriptome sequencing were all located within DNA coding regions. Consequently, we were able to use genomic DNA as a template as this contains both exonic and intronic DNA. Primers were designed according to unigene sequences and applied to genomic DNA templates. Primer binding to a genomic DNA template is known to fail in cases where PCR primers straddle an exon-intron junction (Kim *et al*., 2011). Furthermore, if introns are present within the target amplification, then PCR products were larger in length than the expected and may fail to be amplified. A primer binding site polymorphism in our present template led to inferior or erratic amplification in some cases. Of the 94 consistently amplified primers, 89 loci were polymorphic in *T. granosa* population (Table 1), while only 5 loci were monomorphic. The proportion of polymorphic loci in the present study was 94.68%, which was much higher than blood clam microsatellites that have been previously reported by Shi *et al*. (2008) (58.54%), Zhou *et al*. (2013) (46.77%), and Dong *et al*. (2012, 2013) (76.54%, 73.91%). The number of alleles observed for all the polymorphic loci ranged from 2 to 15, with a mean of 4.84. The observed heterozygosity varied between 0.000 and 1.000, while the expected heterozygosity ranged from 0.102 to 0.921. The mean observed heterozygosity (0.555) across all loci was lower than the expected (0.564). PIC value ranged from 0.048 to 0.886 with a mean of 0.503. The high level of variation of microsatellites length observed during the present study was marginally higher than the reported for blood clam (Shi *et al*., 2008; Zhou *et al*., 2013; Dong *et al*., 2012, 2013). In our study, the proportion of the polymorphic loci was obviously higher than the previous results. Besides, the variation of the genetic parameters, such as N_a, H_o, H_e, PIC, were also higher than those of the previous results in this species. We do not think it is occasional. It may be due to the genetic polymorphism of experimental samples. These samples were collected randomly from Yueqing Gulf, Zhejiang Province. It was a natural population, not an artificial breeding population. So the genetic polymorphism of this population is abundant, which resulted high proportion of polymorphic loci. In total, 32 of the 89 loci deviated significantly from Hardy-Weinberg equilibrium, following Bonferroni correction, and these loci were not suitable for investigation genetic population structure or evaluation genetic resources. The departure from Hardy-Weinberg equilibrium may be attribute to one or more reasons, such as bottleneck effects, the presence of null alleles and heterozygote deficiency at allozyme loci. A previous study of polymorphic microsatellites in *T. granosa* reported that 35 of the 62 loci studied also deviated from Hardy-Weinberg equilibrium (Dong *et al*., 2012). Indeed, deviations from Hardy-Weinberg equilibrium have been demonstrated in a series of studies aiming to identify microsatellite markers for shellfish (Li *et al*., 2011).

Segregation analysis was performed for all 89 polymorphic microsatellites by analyzing one full-sibling family. Data showed that 50 loci were homozygous in the parents (AA×AA or AA×BB), resulting in a single genotype (AA or AB genotype) in the offspring. The remaining 39 loci were polymorphic and segregated within the family, implying that this panel of loci would be very suitable but still deficient for microsatellites-based genetic linkage map studies for *T. granosa*. Genotypic frequencies in parents and offspring at each of these 39 loci are shown in Table 2. As expected, 32 loci were co-dominantly segregated in a Mendelian fashion ($P>0.05$) after Bonferroni correction. Only one locus (*Teg4806*) appeared the null allele. The genotypic ratio of the locus *Teg4806* conformed to Mendelian expectation when we assumed the parent carried a null allele in the heterozygote state, and the offspring genotypes were homozygotes and heterozygotes for the null allele. Null alleles for single-copy, PCR-based DNA markers are often the results of polymorphisms in the microsatellite flanking regions of DNA to which PCR primers are designed to bind (Jones *et al*., 1998). Seven other microsatellites showed significant deviation from expected genotype ratios. The reasons might be as follows: first, strong zygotic selection during the larval and juvenile stages may underlie the cause of such distortion in segregation. A similar phenomenon was put forward for the Pacific oyster family (*Crassostrea gigas*) (Launey *et al*., 2001) and the bay scallop (*Argopecten irradians irradians*) (Li and Li, 2011). Sec-

ondly, when the selection of the favorable allele replacement occurs, under the selection pressure, the corresponding gene frequency will change where the closely linked sites locate with. The allele frequency rises when it is closely linked with favorable selected alleles, and reduces when linked with unfavorable alleles. This represents the result of natural selection as organisms adapt to their environment and eliminate invisible lethal genes. If the microsatellites of unexpected Mendelian segregation ratios were used in a genetic linkage map, they may exert significant effect upon quality (Reece *et al*., 2004). Consequently, there is a very real need to develop a larger number of polymorphic microsatellite markers for further mapping studies.

Table 2 Medelian segregation analysis of 39 microsatellite markers in a full-sibling family of *Tegillarca granosa*

Locus	Sire	Dam	Genotypes of progeny	Expected ratio	Observed ratio	*P*-value
Teg3967	AD	BC	AB:AC:BD:CD	1:1:1:1	2:15:20:11	0.002
Teg4020-1	AB	AB	AA:AB:BB	1:2:1	13:30:7	0.179
Teg4093-2	AA	AB	AA:AB	1:1	19:31	0.090
Teg4121	AB	AA	AA:AB	1:1	18:32	0.048
Teg4170	AA	AB	AA:AB	1:1	32:18	0.048
Teg4227	AB	BB	AB:BB	1:1	16:30	0.039
Teg4303	CD	AB	AC:AD:BC:BD	1:1:1:1	8:5:2:14	0.012
Teg4495	AB	AB	AA:AB:BB	1:2:1	9:22:19	0.094
Teg4516	AA	AB	AA:AB	1:1	35:4	0.000*
Teg4540	AB	BB	AB:BB	1:1	20:30	0.157
Teg4547	AB	AB	AA:AB:BB	1:2:1	16:5:16	0.000*
Teg4613	AA	AB	AA:AB	1:1	25:22	0.662
Teg4654	AB	AB	AA:AB:BB	1:2:1	15:9:9	0.011
Teg4667	AC	AB	AA:AB:AC:BC	1:1:1:1	9:20:6:15	0.025
Teg4719	BD	AC	AB:AD:BC:CD	1:1:1:1	11:6:16:9	0.168
Teg4730	BB	AB	AB:BB	1:1	20:26	0.376
Teg4806	NB	AB	AN:((BB+BN):AB	1:2:1	13:23:10	0.822
Teg4849	BC	AC	AB:AC:BC:CC	1:1:1:1	1:25:15:7	0.000*
Teg4854	AA	AB	AA:AB	1:1	27:19	0.238
Teg4893-2	AB	AA	AA:AB	1:1	27:23	0.572
Teg4996	AA	AB	AA:AB	1:1	27:21	0.386
Teg5082	AD	BC	AB:AC:BD:CD	1:1:1:1	9:10:18:10	0.213
Teg5137	AB	AB	AA:AB:BB	1:2:1	21:13:15	0.002*
Teg5150	AA	AB	AA:AB	1：1	5:45	0.000*
Teg5219	AB	AB	AA:AB:BB	1:2:1	12:21:16	0.438
Teg5262-1	AA	AB	AA:AB	1:1	17:32	0.032
Teg5262-2	BB	AB	AB:BB	1:1	20:26	0.376
Teg5286	AA	AB	AA:AB	1:1	5:45	0.000*
Teg5416	AC	BC	AB:AC:BC:CC	1:1:1:1	12:14:12:10	0.881
Teg5440	AA	AB	AA:AB	1:1	18:31	0.063
Teg5468	AB	AA	AA:AB	1:1	16:25	0.160
Teg5478	AB	AA	AA:AB	1:1	24:26	0.777
Teg5627	AB	AA	AA:AB	1:1	28:22	0.396
Teg5790	AB	AB	AA:AB:BB	1:2:1	5:23:20	0.009
Teg5791	AB	AB	AA:AB:BB	1:2:1	10:29:8	0.254
Teg5870	BB	AB	AB:BB	1:1	27:21	0.386
Teg6682	AB	AA	AA:AB	1:1	23:24	0.884
Teg6790	AA	AB	AA:AB	1:1	28:19	0.189
Teg6841	AB	AB	AA:AB:BB	1:2:1	10:16:23	0.002*

Notes: * *P*-value, Significant deviation ($P<0.05$) from expected Mendelian ratios after Bonferroni correction ($k=39$). N, represent inferred null alleles.

Given the level of polymorphism observed in population studied, 57 of the 89 polymorphic markers are highly likely to be useful in the analysis of population structure and genetic diversity in species where resource investigations, conservation, and management programs are important. Furthermore, the 32 microsatellites identified co-dominantly segregated in a Mendelian fashion may be uniquely useful in constructing genetic linkage maps and assisting the development of QTL mapping.

4 Conclusion

Our transcriptome database of *T. granosa* contains a large number of sequence information, which is very suitable for the rapid and large-scale development of microsatellite markers. With the development and improvement of the identification of molecular technique, more and more microsatellite markers will be obtained.

There are 89 novel microsatellites identified in the current study, which will greatly enrich the microsatellite markers resource data of *T. granosa*. As a powerful molecular tool, microsatellites can be applied for further studies of biodiversity analysis, phylogeny, and molecular marker-assisted selection.

Acknowledgements

The study was supported by the National High Technology Research and Development Program of China (863 Program) (No. 2012AA10A410), the National Project for Agricultural Technology System (No. CARS-48), National Infrastructure of Fishery Germplasm Resources of China (No. 2015DKA30470), and Zhejiang Major Program of Science and Technology (No. 2012C12907-4).

References

Abdelkrim, J., Robertson, B. C., Stanton, J. L., and Gemmell, N. J., 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques*, **46** (3): 185-192.

Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., and Rysgaard, S., 2012. Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *The ISME Journal*, **6**: 11-20.

Broom, M. J., 1983. Gonad development and spawning in *Anadara granosa* (L.) (Bivalvia: Arcidae). *Aquaculture*, **30**: 211-219.

Castoe, T. A., Polle, A. W., Gu, W., de Koning, A. P. J., Daza, J. M., Smith, E. N., and Pollock, D. D., 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources*, **10**: 341-347.

Dong, Y. H., Wu, G. X., Yao, H. H., Bian, P. J., Zhao, D. M., and Lin, Z. H., 2013. Characterization of 34 polymorphic EST-SSR markers in *Tegillarca granosa* and their transferability in *Anadara craticulata*. *Journal of Fisheries of China*, **37**: 70-77.

Dong, Y. H., Yao, H. H., Lin, Z. H., Zhang, L. L., and Bao, Z. M., 2012. Characterization of 62 polymorphic EST-SSR markers in the blood clam (*Tegillarca granosa*) and their cross-amplification in *Scapharca subcrenata*. *Conservation Genetic Resources*, **4**: 991-997.

Gu, X. Y., Zeng, Q. G., You, Z. J., and Lin, Z. H., 2008. Isolation and characteristics of six microsatellite primers of *Tegillarca granosa*. *Oceanologia et Limnologia Sinica*, **39**: 661-664.

Hubert, S., and Hedgecock, D., 2004. Linkage maps of microsatellite DNA marker for the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research*, **23**: 379-385.

Jones, A. G., Stockwell, C. A., Walker, D., and Avise, J. C., 1998. The molecular basis of a microsatellite null allele from the white sands pupfish. *Journal of Heredity*, **89**: 339-342.

Kim, W. J., Jung, H., and Gaffney, P. M., 2011. Development of type I genetic markers from expressed sequence tags in highly polymorphic species. *Marine Biotechnology*, **13**: 127-132.

Li, Q., Park, C., and Kijima, A., 2002. Isolation and characterization of microsatellite loci in the Pacific abalone, *Haliotis discus hannai*. *Journal of Shellfish Research*, **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 in the Pacific oyster (*Crassostrea gigas*). *Genes & Genomics*, **33**: 313-316.

Li, R. H., and Li, Q., 2011. Mating systems and reproductive success in hermaphroditic bay scallop, *Argopecten irradians irradians* (Lamarck 1819), inferred by microsatellite-based parentage analysis. *Journal of the World Aquaculture Society*, **42**: 888-898.

Liu, B., Teng, S. S., Shao, Y. Q., Chai, X. L., and Xiao, G. Q., 2012. Isolation and characterization of 39 novel polymorphic EST-SSR loci for the blood clam, *Tegillarca granosa*. *Conservation Genetic Resources*, **4**: 375-378.

Liu, Y. G., Chen, S. L., and Li, B. F., 2004. Assessing the genetic structure of three Japanese flounder (*Paralichthys olivaceus*) stocks by microsatellite markers. *Aquaculture*, **243**: 103-111.

Macas, J., Neumann, P., and Navratilova, A., 2007. Repetitive DNA in the pea (*Pisum sativum L.*) genome: Comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. *BMC Genomics*, **8**: 427.

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 unguivulata* (L) Walp). *BMC Genomics*, **12**: 8.

Rahmann, S., Beygo, J., Kanber, D., Martin, M., Horsthemke, B., and Buiting, K., 2013. Amplikyzer: Automated methylation analysis of amplicons from bisulfite flowgram sequencing. *Peer J Preprints*, http://peedj.com/preprints/122v2.

Reece, K. S., Ribeiro, W. L., Gaffney, P. M., Carnegie, R. B., and Allen, S. K., 2004. Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): Confirmation of null alleles and non-Mendenlian segregation ratios. *Journal of Heredity*, **95**: 346-352.

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

Shi, S. F., Yao, H. H., Lin, Z. H., Zhou, X. L., Qi, X. Y., and Dong, Y. H., 2013. Characterization and analysis of 24 polymorphic EST-SSR loci in *Tegillarca granosa*. *Marine Sciences*, **37** (8): 42-46.

Sithichoke, T., Prakit, S., pichahpuk, U., Juntima, C., Duangjai, S., Worapa, S., Warunee, S., Somvong, T., and Peerasak, S., 2009. Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiate* (L.) Wilczek). *BMC Plant Biology*, **9**: 137.

Wang, H. X., Huan, P., Lu, X., and Liu, B. Z., 2011. Mining of EST-SSR markers in clam Meretrix meretrix larvae from 454 shotgun transcriptome. *Genes & Genetic Systems*, **86**: 197-205.

Wang, R. C., Wang, Z. P., and Zhang, J., 2008. *Science of Marine Shellfish Culture*. Ocean University of China Press, Qingdao, 337-339 (in Chinese).

Wicker, T., Schlagenhauf, E., Graner, A., Close, T. J., Keller, B., and Stein, N., 2006. 454 sequencing put to the test using the complex genome of barley. *BMC Genomics*, **7**: 275.

Xiong, G., Ma, X., Wang, X. Q., Zeng, Z. N., Zhu, D. L., Kang, L., and Wu, Q. S., 2015. Development and characterization of 18 microsatellite markers for the *Babylonia lutosa*. *Conservation Genetic Resources*, **7**: 471-472.

Zhou, X. L., Zhu, J. H., Dong, Y. H., Lin, Z. H., Fu, K. L., Yao, H. H., and Qi, X. Y., 2013. Development and comparative study of genomic-SSR and EST-SSR in *Tegillarca granosa*. *Oceanologia et Limnologia Sinica*, **44** (2): 467-475.

(Edited by Qiu Yantao)