© Indian Academy of Sciences

ONLINE RESOURCES

Identification of forty-five gene-derived polymorphic microsatellite loci for the sea cucumber, *Apostichopus japonicus*

MENG CHEN[1], LINLIN GAO[1], WEIJIE ZHANG[1], HONGZHENG YOU[2], QIAN SUN[1] and YAQING CHANG[1*]

[1]*Key Laboratory of Mariculture, DaLian Ocean University, Liaoning 116023, People's Republic of China*
[2]*Tianjin Fishery Technical Extension Station, Tianjin 300221, People's Republic of China*

[Chen M., Gao L., Zhang W., You H., Sun Q. and Chang Y. 2013 Identification of forty-five gene-derived polymorphic microsatellite loci for the sea cucumber, *Apostichopus japonicus*. *J. Genet.* **92**, e31–e35. Online only: http://www.ias.as.in/jgenet/OnlineResources/92/e31.pdf]

Introduction

Sea cucumber (*Apostichopus japonicus*) is a common benthic detritus feeder of coastal sea habitats of East Asia, including Japan, China, Korea and fareastern Russia (Chang *et al.* 2009). Because of its high nutritive value and use in traditional medicine, sea cucumber is considered as one of the most valuable species in aquaculture markets in China. However, the number of wild sea cucumbers has severely declined owing to overexploitation and environment deterioration (Peng *et al.* 2009). To increase production of sea cucumber, artificial breeding techniques were established in China in the 1980s and sea cucumber aquaculture has been developing rapidly in recent years (Qiao and Cheng 2005).

Because sea cucumber is an important aquaculture species in China, application of marker-assisted selection (MAS) or genomewide marker-assisted selection (G-MAS) in the *A. japonicus* breeding programme is expected to be a fertile research area. To apply MAS and G-MAS, a saturated linkage map should be constructed first, using large numbers of molecular markers. Unfortunately, progress in genetic and genomic research in sea cucumber has relatively slower than in other economically important marine species (Cecilia *et al*. 2010; Guo *et al*. 2011; Wang *et al*. 2011; Zhang *et al.* 2011). There is only one preliminary genetic map based on 37 amplified fragment length polymorphism (AFLP) primer combinations and 20 microsatellite (simple sequences repeat; SSR) markers from 88 individuals (Li *et al.* 2009). While this framework linkage map is useful, the marker density is still low. Mapping additional markers, especially appropriate codominant markers, is particularly important. Some microsatellite markers (Kanno *et al*. 2005; Zhan *et al*. 2007; Peng *et al*. 2009; Liao *et al*. 2011) and single-nucleotide polymorphism (SNPs) (Sun *et al*. 2010; Yang *et al*. 2012) have been developed for sea cucumber. However, the number of markers is still insufficient for planned QTL analysis of traits, such as growth or disease resistance. More codominant markers on the maps are desirable.

Compared with type II microsatellite markers associated with anonymous genomic sequences, type I markers, associated with genes of known function, are more useful for genome mapping because they serve as anchorage points for genomic segments for which genomic syntenies have been analysed (Serapion *et al*. 2004). Additionally, type I markers are associated with genes that are conserved in a wide spectrum of species through evolution, allowing comparative genome analysis and study of genome evolution. In contrast, type II markers, which are associated with anonymous sequences, can only occasionally be identified among very closely related species, allowing comparative genome analysis of these species. Despite their high utility, type I markers are not easy to identify, because that usually requires founding and sequencing of microsatellite-enriched cDNA libraries. The whole process used to be time-intensive and cost-intensive especially for organisms that are not model organisms, for which genomic information is limited. Progress in sequencing technologies has enabled generation of gigabases of DNA sequence in a short time and at minimal cost (Davey and Blaxter 2010), which makes it possible to discover type I markers in an efficient and economical way.

In this study, we identified 45 gene-associated microsatellite markers from a cDNA library constructed by transcriptome sequencing for linkage mapping, comparative genome analysis and genetic breeding in *A. japonicus*.

Materials and methods

Transcriptome sequencing

A transcriptome of *A. japonicas* was sequenced in our laboratory using the Illumina HiSeq 2000 Sequencing Technology.

*For correspondence. E-mail: yaqingchang@hotmail.com.

Keywords. microsatellite markers; transcriptome; EST; sea cucumber.

To maximize transcript representation for high-throughput sequencing, different adult tissues were collected, including intestine, respiratory tree and longitudinal muscle. Total RNA was extracted from each tissue using TRIzol ® reagent (Invitrogen, Carlsbad, USA) following the manufacturer's protocol. Equal quantities of high-quality RNA were pooled for cDNA synthesis. After cDNA samples were prepared, approximately 4 μg of the mixed cDNA pool was used for sequencing using Illumina HiSeq 2000. More than 12,000 UniGene sequences, representing a particular continuous path through a set of contigs, were obtained. These sequences represent transcribed sequences of the genome, or genes. For gene annotation, each UniGene was compared against the Swiss-Prot (http://www.expasy.ch/sprot/) database using BlastX with an E-value threshold of 1e-6. The top 20 hits extracted from the BlastX results were used for gene annotation. Gene ontology (GO; http://www.geneontology.org) and KEGG pathway (Kyoto Encyclopedia of Genes and Genomics; http://www.genome.jp/kegg/pathway.html) analysis recovered diverse biological functions and processes.

Primer design

The SSRs were obtained by screening through these 12,000+ UniGene sequences using SSRHunter 1.3 software (http://www.bio-soft.net/dna/SSRHunter.htm) with parameters of at least six repeats for dinucleotides, five repeats for trinucleotides and four for tetranucleotides. After each search, a total 355 microsatellite length and position data pairs within the sequence were recorded. Among them, 110 sequences with appropriate flanking sequences and GC content were selected for PCR primer design (Primer Premier ver. 5.0 software; http://www.PremierBiosoft.com/faq.html). All these 110 primer pairs were used in this study for marker identification, and were synthesized by the Sangon, Shanghai, China.

DNA extraction, PCR amplification and genotyping

Fifty individuals of sea cucumber were collected from Zhangzidao island in Liaoning province, China. Genomic DNA was extracted from longitudinal muscle using Marine Animals DNA kit (Qiagen, Hilden, Germany). The extracted genomic DNA was stored at −20°C until genotyping. PCR amplification was carried out in 15-μL reaction volumes containing 40–50 ng of template DNA, 0.6 μM of each primer, 7.5 μL 2× Taq PCR MasterMix (Qiagen; 0.1 U of Taq DNA polymerase/μL, 500 μM dNTP each, 20 mM Tris–HCl, 100 mM KCL, 3 mM $MgCl_2$). Amplifications were done using Eppendorf Mastercycler ep gradient S thermal cycler (Eppendorf, Hamburg, Germany). PCR amplification protocol involved initial denaturing for 5 min at 94°C, followed by 26 cycles of 30 s denaturation at 94°C, annealing at the temperature given in table 1 for 30 s, extension of 30 s at 72°C, and a prolonged extension step of 10 min at 72°C. PCR products were analysed using electrophoresis on 6% denaturing polyacrylamide gels in 1× TBE buffer and visualized with silver staining. Images were scanned with an Epson Perfection 2480 scanner and analyzed using Gel-Pro Analyzer (ver. 4.5) (http://www.mediacy.com/index.aspx?page=GelPro). Sizes of amplified fragments were determined by reference to a standard base-pair ladder, 20-bp DNA ladder marker (TaKaRa, Dalian, China).

Genetic data analysis

Number of alleles (N_a), and observed (H_o) and expected (H_e) heterozygosity were calculated using PopGene ver. 3.2 (Yeh *et al*. 2000). Significance of scoring error due to stuttering, large-allele drop-out and null alleles at 95 % confidence interval were evaluated using MicroChecker software (Van Oosterhout *et al*. 2004). Tests for linkage disequilibrium and deviations from Hardy–Weinberg equilibrium (HWE) were performed using Genepop ver. 4.0 (Rousset 2008). Significance criteria were adjusted for the number of simultaneous tests using Bonferroni correction (Rice 1989).

Results and discussion

Forty-five of the 110 primer pairs used revealed genetic polymorphism in the population of *A. japonicus* (table 1). Number of alleles per polymorphic locus ranged from 2 to 20 which gives an average of 5.53 alleles per locus. In general, the present result is lower than the numbers in previous studies, which looked at noncoding regions (Kanno *et al*. 2005; Zhan *et al*. 2007; Liao *et al.* 2011), because gene sequences tend to be less polymorphic due to functional constraints. While noncoding sequences are relatively less constrained and show higher levels of polymorphism, sequences within protein-coding regions generally show lower mutation rates because of evolutionary conservation (Serapion *et al.* 2004).

Observed and expected heterozygosity varied from 0.1333 to 0.9000 and from 0.1266 to 0.9463, which give averages of 0.4486 and 0.6579, respectively. Significant departures from HWE were observed in 17 of the 45 single-locus exact tests after sequential Bonferroni correction. Some primer pairs could not amplify any product in a few individuals (e.g. at loci AJ14707, AJ66748 and AJ78245). Repeated DNA extraction and PCR did not result in any improvement in PCR product synthesis in these individuals at these loci, indicating high probability of null alleles. MicroChecker analysis suggested that there were no indications for scoring error due to stuttering and large-allele drop-out ($P < 0.05$). Null alleles in some individuals at AJ14707, AJ66748 and AJ78245 may be plausible reason for the observed deviation from HWE for these loci. Previous studies also came up with null alleles as a common phenomenon at microsatellite loci in *A. japonicus* (Kanno *et al*. 2005; Zhan *et al*. 2007; Peng *et al*. 2009; Liao *et al*. 2011). No significant linkage disequilibrium was detected between any pair of loci.

Table 1. Characterization of 45 microsatellite loci in sea cucumber (*Apostichopus japonicus*).

Locus ID	Primer sequences (5'-3')	Repeat motif	Size (bp)	T_a (°C)	N_a	H_o	H_e	P_{HWE}
AJ01958	F:GAACGTTAGGGTGTGGCATT R:TGGTGTAGTGGTGAGGGTGA	$(CCT)_6$	157–215	61	10	0.6667	0.8689	0.0838
AJ04841	F:AACGCCGACGATATGTCAAT R:TTGGGATGCTTGTCTCAACA	$(TAA)_5$	184–225	60	6	0.4286	0.7221	0.1258
AJ07630	F:ATGACGCTAACGCTGACGAT R:GCATAACCTCACTGCACGTC	$(CGA)_5$	135–209	60	5	0.5417	0.5089	0.3264
AJ08658	F:AAGGAGCCACTACGATGACG R:ACCGTCAAAACCCACAAAAA	$(TGT)_5$	175–215	60	9	0.7586	0.8675	0.0708
AJ09045	F:TTCCAGTTTTATGGGGCTGT R:TGGGAGGGTAACACACACAA	$(AAG)_6$	170–254	60	5	0.2759	0.3575	0.1460
AJ10026	F:GATGGTGAAGACGAAGAGGA R:TCCATGAATTCCTCGTTATCAA	$(GAA)_6$	185–209	60	5	0.8667	0.6469	0.0287
AJ10675	F:AGATGTCAGCCACATGCAAC R:GGTAAGCTTGTGGGAATGGA	$(TAG)_5$	185–190	60	2	0.3103	0.2668	0.3545
AJ13361	F:TGGAAGACGAAGATGAGCAA R;GGAATGACCCTACGTCCAAA	$(CAG)_5$	165–205	60	2	0.1333	0.1266	0.7367
AJ14707	F:TCGAGGTTGTCCCTGCTATC R:GCTGAGTTGATTGCACAAGG	$(GAG)_7$	170–201	60	6	0.1923	0.6983	0.0000*
AJ16430	F:TGGCTCATCCGTTGATTACA R:AATCGGTGCAAGGTAGCTGA	$(CAA)_5$	162–172	63	3	0.2500	0.6045	0.0000*
AJ17541	F:CAAACCAGGCTCAAATTCCA R:GGTTCTTTGTAACATGTCGGATT	$(AAC)_5$	195–210	60	5	0.3793	0.6673	0.0000*
AJ18534	F:AGAAAAGCGGTCATGGACAC R: CGGCAATCTGAATCTTCTCC	$(AG)_6$	160–175	63	4	0.2857	0.5890	0.0000*
AJ19024	F:TCGTGGAAAAAGTTGGATTTG R:GCTGCTATCGGTCTCTTTCG	$(GAA)_{19}$	163–280	61	20	0.4667	0.9463	0.0994
AJ19270	F:CCCGATATCGTCACCTTCAT R:CGACAATCCCTGAAAAGGAA	$(TCC)_9$	220–250	60	6	0.4074	0.6045	0.0740
AJ19640	CATCCTCCCAACAACCTCAG GCTGTCGAAGCATGGGATAC	$(CAG)_6$	140–183	61	7	0.3448	0.6721	0.0960
AJ19701	F:TTGAAACGGAAGAGCCAGTT R:TGTTTGGTTGCTTTGGTTGA	$(AAG)_7$	145–184	60	5	0.6538	0.7044	0.1138
AJ20333	F:CGTCGACCAAAGAGAGCAAT R:TCCCTGGACTGGCACTAATC	$(AGT)_7$	181–198	60	4	0.6154	0.598	0.4212
AJ20385	F:AGCAAACCACCGAGTACACC R:CTCCACCACTCTCCGATTCT	$(CAA)_5$	187–205	60	4	0.4000	0.5768	0.0216
AJ20674	F:GTCCAACATGAAGCCTGGTT R:GCACCGTACCCTAGACCTTG	$(CAA)_9$	157–202	61	8	0.6071	0.7162	0.9780
AJ20700	F: ATGAAGACACCCCAGATTGC R: TTGTCAGAAGGTTGGATGAGG	$(AG)_6$	192–205	62	4	0.1453	0.4847	0.0861
AJ21199	F:TACGCCTTTTGTCCGTTTTC R:TGCAAGGCACAATTCTAAAAGA	$(CTA)_6$	162–216	61	4	0.5000	0.474	0.2220
AJ21274	F:GAGGGAAGAGACCACCGATA R:AACCTCTCGAGTCATTCTCCA	$(AG)_{10}$	175–206	64	7	0.8668	0.7695	0.1895
AJ21455	F:CAGCGGAGACAAAGCAGATA R:GCCGTATCATTACATGCGAGA	$(CT)_7$	178–235	61	8	0.6250	0.8298	0.1992
AJ21478	F:AACCTTGAAATGCGACGAAG R:ACCTGCCTGGTAGACCACAG	$(TGA)_5$	148–190	61	7	0.2000	0.7373	0.0000*
AJ22011	F:GCGGCCTGTTTCTTAGGATT R:GCCTCATTACTGCCCATGAT	$(AT)_7$	170–202	60	4	0.7000	0.5186	0.2225
AJ22049	F:CGGGACACCTACGAAAGGTA R:CATACCCCATCTCCTGCACT	$(AAGG)_4$	178–202	60	5	0.7931	0.7520	0.0110
AJ51083	F: CGGAAAAGAGGAAGAGAAAGG R: GCTCCTTCGTCATGTACTCCA	$(ATC)_5$	165–185	60	5	0.3333	0.7446	0.0000*
AJ52273	F:CGACGTCGAATCAAACAAAG R:CAGCTCTTTCCTCCCTTCCT	$(AAG)_6(AGA)_5$	113–194	58	8	0.4800	0.7935	0.1106
AJ54445	F:TGAGGGAGAGCAAGAAGTGG R:TCTCTCCATCACTGCCCTCT	$(GAG)_6$	190–220	58	5	0.2593	0.6918	0.0000*
AJ64884	F: CGCTGTTTAATTCCTCATTGTG R: GCTTAAACACGCGCAAAAA	$(GT)_6$	150–173	60	3	0.1714	0.6526	0.0000*

Table 1 (*contd.*)

Locus ID	Primer sequences (5'-3')	Repeat motif	Size (bp)	T_a (°C)	N_a	H_o	H_e	P_{HWE}
AJ66748	F:CCTTAACGGAGCAGTTGGAG R:CCGTCTCCACCAAGATCAAT	$(GAG)_6$	189–204	60	4	0.3200	0.7159	0.0050*
AJ72502	F:TGGGAAATCCTCCTCTTCCT R:TCGCTGCCTGAACTATCTGA	$(AGC)_5$	150–224	60	4	0.3667	0.7102	0.0000*
AJ78166	F:GGCTGCATCGTCTTTAGCAT R:TCCCATACAGCAAGAGTAGCC	$(TCA)_5$	142–163	60	5	0.7333	0.7006	0.0038*
AJ78245	F:CTTGATGTCCTGCTGTCGAA R:CTACCACGCCCATTTTGACT	$(TCT)_5$	190–210	61	6	0.4231	0.8296	0.0000*
AJ81104	F:GGGAGGGGGAAGATCAGATA R:TCCTCATCCGGTTTATCCAC	$(AGA)_5$	221–233	60	3	0.2333	0.5588	0.0000*
AJ84803	F:TCGGTTGTTGTTTTGCATGT R:AAGCAGAGGAACAGGACAGG	$(CCT)_5$	177–193	60	4	0.3103	0.6830	0.0000*
AJ87449	F:AATCTGTTTCCGGTGACTCG R:CCCACCTTCATTCTCCTCCT	$(AGA)_5$	142–160	60	4	0.6296	0.724	0.0333
AJ87669	F:TCCAAATCCACCACCATTCT R:ATTTTGGAGGTCATGGCAAC	$(TCC)_5$	174–192	60	4	0.4000	0.7215	0.0002*
AJ89887	F:CGATGATGAGCTGCAAATGT R:GGGTGACAAGGAGTCTCCAA	$(AGG)_5$	153–174	60	5	0.5172	0.6062	0.3359
AJ90649	F:TGCCATAGCTGGAGCTAACA R:ATGATGTGGATGGGGAAGAA	$(CCT)_7$	165–180	61	4	0.1481	0.7233	0.0000*
AJ91092	F: AGACATACACGGTGGCCAAT R: GTTGGCCAAACTCAAAAGGA	$(ACT)_5$	211–228	60	4	0.1667	0.4831	0.0000*
AJ91208	F:TCCGGTTTAGAAGAGGAGCA R:AGGGTACAGGATGCCTCGTT	$(CGTT)_4$	155–204	60	8	0.4000	0.7073	0.0660
AJ91727	F:TCGGTTATGGGTTCGTTAGC R:TTTTCCAACTCATTACGATTTTCA	$(TA)_8$	196–224	60	6	0.9000	0.6825	0.1258
AJ91732	F: TGCAAGGAGTGTACCACAATG R: GCCAACCAGACTATGGAAGC	$(AT)_6$	226–255	60	6	0.4815	0.8225	0.0584
AJ91905	F:GATCCAGAAGTACCCCACCA R:CAATTTGAGGAGGGTGGAAA	$(TA)_7$	160–196	61	6	0.5000	0.7462	0.0839

T_a, annealing temperature; N_a, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; P_{HWE}, possibilities to fit Hardy–Weinberg equilibrium using χ^2tests. *Not in conformity with Hardy–Weinberg equilibrium ($P < 0.01$, after Bonferroni correction).

In conclusion, 45 novel gene-derived markers were identified from transcriptome sequences. We hope these markers will be useful for genetic linkage mapping, comparative genome analysis and genetic breeding in *A. japonicas*.

Acknowledgements

This research work was supported by The Key Project of the Modern Agricultural Field ('863' Program) of China (2012AA10A412), and The Key Scientific Specific Project of Liaoning Province , China (2011203003).

References

Cecilia C. S., Kanako F., Akiyuki O., Osamu H., Takashi S., Kagayaki M. *et al.* 2010 A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *BMC Genomics* **11**, 554.

Chang Y. Q., Feng Z. G., Yu J. P. and Ding J. 2009 Genetic variability analysis in five populations of the sea cucumber *Stichopus (Apostichopus) japonicus* from China, Russia, South Korea and Japan as revealed by microsatellite markers. *Mar. Ecol.* **30**, 455–461.

Davey J. L. and Blaxter M. W. 2010 RADSeq: next-generation population genetics. *Brief Funct. Genomics* **9**, 416–423.

Guo X., Li Q., Wang Q. Z. and Kong L. F. 2011 Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. *Mar. Biotechnol.* **14**, 218–226.

Kanno M., Li Q. and Kijima A. 2005 Isolation and characterization of twenty microsatellite loci in Japanese sea cucumber (*Stichopus japonicus*). *Mar. Biotechnol.* **7**, 179–183.

Li Q., Chen L. and Kong L. 2009 A genetic linkage map of the sea cucumber, *Apostichopus japonicus* (*Selenka*), based on AFLP and microsatellite markers. *Anim. Genet.* **40**, 678–685.

Liao M. J., Wang Y. G., Rong X. J., Zhang Z., Li B., Wang L. *et al.* 2011 Development of new microsatellite DNA markers from *Apostichopus japonicus* and their cross-species application in *Parastichopus parvimensis* and *Pathallus mollis*. *Int. J. Mol. Sci.* **12**, 5862–5870.

Peng W., Bao Z. M., Du H. X., Dong Y., Lv Y. M, Zhou Z. C. and Hu J. J. 2009 Development and characterization of 38 novel EST-SSRs for the sea cucumber *Apostichopus japonicas*. *Conserv. Genet. Resour.* **1**, 447–450.

Qiao J. and Cheng B. 2005 The situation and prospect of pond culture for *Apostichopus japonicus*. *Mar. Sci.* **29**, 80–82 (in Chinese).

Rice W. R. 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. Res.* **8**, 103–106.

Serapion J., Kucuktas H., Feng J. and Liu Z. 2004 Bioinformatic mining of type I microsatellites from expressed sequence tags of Channel catfish (*Ictalurus punctatus*). *Mar. Biotechnol.* **6**, 364–377.

Sun W. J., Li Q. and Kong L. F. 2010 Characterization of thirteen single nucleotide polymorphism markers in the sea cucumber (*Apostichopus japonicus*). *Conserv. Genet. Resour.* **2**, 141–144.

Van Oosterhout C., Hutchinson W. F., Wills D. P. M. and Shipley P. 2004 Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538.

Wang C. M., Bai Z. Y., He X. P., Lin G., Xia J. H. , Sun F. *et al.* 2011 A high-resolution linkage map for comparative genome analysis and QTL fine mapping in Asian seabass, *Lates calcarifer*. *BMC Genomics* **12**, 174.

Yang A., Sun D., Liu S., Dong Y., Chen Z. and Zhou Z. 2012 Characterization of fifteen SNP markers by mining EST in sea cucumber, *Apostichopus japonicus. J. Genet.* **91**, e49–e53.

Yeh F. C., Yang R., Boyle T. J., Ye Z. and Xiyan J. M. 2000 PopGene32, Microsoft Windows-based freeware for population genetic analysis. Molecular Biology and Biotechnology Centre. University of Alberta, Edmonton, Canada.

Zhan A. B., Bao Z. M., Lu W., Hu X. L., Peng W., Wang M. L. *et al.* 2007 Development and characterization of 45 markers for sea cucumber (*Apostichopus japonicus*). *Mol. Ecol. Notes* **7**, 1345–1348.

Zhang Y., Xu P., Lu C., Kuang Y., Zhang X., Cao D. *et al.* 2011 Genetic linkage mapping and analysis of muscle fiber-related QTLs in common crap (*Cyprinus carpio L.*). *Mar. Biotechnol.* **13**, 376–392.

Received 22 August 2012, in the revised form 8 October 2012; accepted 10 October 2012
Published on the Web: 20 May 2013