Journal of Genetics, Vol. 97, Suppl 1, December 2018, Online Resources, pp. e61–e65
https://doi.org/10.1007/s12041-018-0935-7

ONLINE RESOURCES

Identification and characterization of 23 microsatellite loci for *Chlorostoma rustica* based on RAD-seq

LING WANG[1], HONG YU[1]* and QI LI[1,2]

[1]*Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, People's Republic of China*
[2]*Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, People's Republic of China*
*For correspondence. E-mail: hongyu@ouc.edu.cn.

Received 26 July 2017; revised 24 September 2017; accepted 3 October 2017; published online 6 June 2018

Abstract. *Chlorostoma rustica* is an ecologically and economically important species in China, which plays an important role on the overall function of the coastal systems. Understanding of the genetic structure of *C. rustica* populations is vital to breeding strategies and conservation programmes. In this study, we isolated and characterized 23 microsatellite loci with high polymorphism using the restriction site-associated DNA sequencing approach. The number of alleles per locus ranged from 6 to 18, and the expected heterozygosities varied from 0. 760 to 0.936. All the polymorphism information content values of the 23 loci were greater than 0.5, indicating that these markers were highly informative and laid the foundation for further genetic analysis.

Keywords. RAD-seq; microsatellite loci; *Chlorostoma rustica*.

Introduction

Chlorostoma rustica is economically important mollusk in China and is naturally distributed along the coasts of China, Japan and Korea (Wang and Wei 1994). *C. rustica* is one of the most valuable and important fishery resources in China. Because of its taste and high nutritional value, *C. rustica* is well-liked by consumers. However, owing to the ever increasing demand for *C. rustica* in domestic markets, overexploitation and habitat destruction, the natural resources of *C. rustica* have declined dramatically in the past decades. Overfishing was also considered as a major reason in the depletion of many marine species (Hutchings 2000; Jackson *et al.* 2001). Thus, to manage and protect the genetic resources of *C. rustica*, studies on its genetic diversity and population structure are necessary. Microsatellites are of high polymorphism, neutrality, codominance and genomewide abundance, becoming one of the most widely used molecular markers in population genetic studies (Morgante *et al.* 2002). Now, with the advent of next-generation sequencing, microsatellite markers can be developed quickly and at relatively low-cost (Berman *et al.* 2014). In this study, we applied the restriction site-associated DNA sequencing (RAD-seq) approach to develop microsatellite markers for *C. rustica*, which will be useful for assessing genetic diversity and structure in populations, and the effective conservation and management of *C. rustica*.

Materials and methods

Genomic DNA was extracted from the foot of *C. rustica* which was collected from Dalian, Liaoning, China for RAD-seq. The RAD-seq library was constructed according to the protocol reported by Baird *et al.* (2008). Briefly, genomic DNA was digested with *Eco*RI and then heat-inactivated at 65°C. The P1 adapter containing individual-specific nucleotide barcodes was ligated to the digested product, which was then randomly sheared and size-selected. dATP overhangs were added to the DNA fraction. DNA was then ligated to a second adapter (P2), which contains T overhangs. The ligated material was purified, eluted and subjected to PCR enrichment. Sequencing was performed on an Illumina HiSeq2500 with pair-end 150-bp reads. Raw reads were processed to get high quality

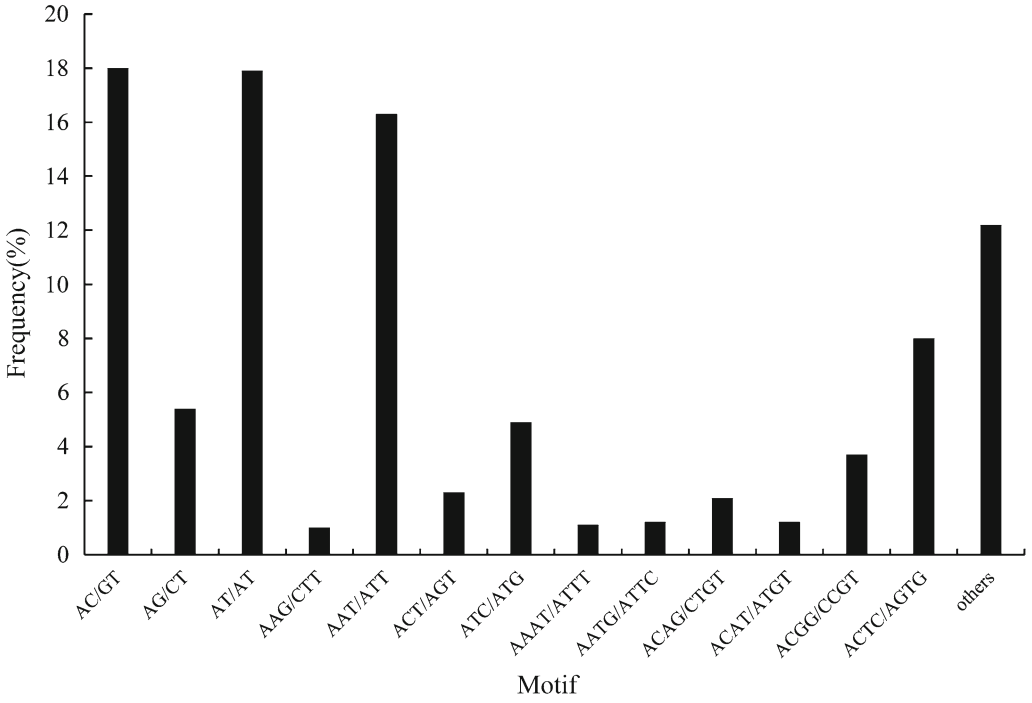

Figure 1. Distribution of microsatellite type in *C. rustica*.

clean reads according to stringent filtering standards: (i) removing reads with ≥10% unidentified nucleotides (N); (ii) removing reads with >50% bases having phred quality scores of ≤20; and (iii) removing reads aligned to the barcode adapter. We filtered out Illumina short reads lacking sample-specific barcodes and expected restriction enzyme motifs before reads clustering. All the short reads from the sample were then clustered into tag reads on the basis of sequence similarity of the first read. The paired-end reads associated with each RAD cluster tag were extracted to construct scaffolds using adjacent contigs identified by paired-end information. Subsequently, the microsatellite mining was performed by using microsatellite identification tool (Thiel *et al.* 2003) and the primer pairs for each microsatellite locus were designed with Primer3 v2.3.6 (http://primer3.sourceforge.net). Primers were designed according to the following criteria: with the size of PCR products ranging from 100 to 300 bp, primer annealing temperature range of 55.0–62.0°C and GC content in the range of 45–60%. Eighty primer pairs were tested for amplification and polymorphism using DNA of 48 *C. rustica* individuals collected from Dalian, Liaoning, China. PCRs were performed in a volume of 10 μL containing 0.25 U *Taq* DNA polymerase (Takara), 1× PCR buffer (Mg^{2+} plus), 0.2 mM dNTP mix, 1 mM of each primer set and 50 ng of genomic DNA. The PCR amplification conditions were as follows: initial denaturation at 94°C for 3 min, 35 cycles of 94°C for 30 s, 30 s at the optimal annealing temperature and 72°C for 1 min, final extension at 72°C for 5 min. Amplification products were separated on 6% denaturing polyacrylamide gel, we used a 10-bp DNA ladder (Invitrogen, Carlsbad, USA) to determine allele sizes. In the analysis of microsatellite loci, the number of alleles (N), observed heterozygosities (H_o), expected heterozygosities (H_e) were estimated using the program Microsatellite Analyser (Dieringer and Schlötterer 2003). Tests for linkage disequilibrium (LD) between all possible loci combinations and deviations from Hardy–Weinberg equilibrium (HWE) were conducted using GenePop v4.0 (Rousset 2008). The polymorphism information content (PIC) was calculated using Cervus v.3 (Kalinowski *et al.* 2007).

Results and discussion

Here, we demonstrated the utility of RAD-seq technique for isolating microsatellite loci in *C. rustica*, and about 4.473 G bases of clean reads were obtained. *De novo* assembly generated 634,266 high-quality contigs with an average size of 290 bp (N50 = 304). Total number of identified microsatellites were 18,767, including 8141 dinucleotide repeats, 5386 trinucleotide repeats, 4201 tetranucleotide repeats, 509 pentanucleotide repeats and 530 hexanucleotide repeats (figure 1). Among these microsatellite loci of *C. rustica*, dinucleotide repeats were the most common microsatellite motif (41.3%) and AC/TG type repeats had the largest proportion (18%). The results confirmed that the RAD-seq approach was efficient for

Table 1. Characterization of 23 microsatellite loci isolated from *C. rustica*.

Locus	Primer sequence (5′–3′)	Repeat motif	No. of alleles	Size range (bp)	T_a (°C)	H_o	H_e	*P* value	PIC	GenBank accession number
Xtw04	F: CATGTTTTCTTGCTTTTTCACG R: AACTGGTTGGCCTGGAGTTA	$(TG)_{23}$	11	270–320	60	0.744	0.851	0.117	0.822	MF460397
Xtw08	F:GCTGGGACTGCTGAGAGACT R:TTGGTAGTTGTTGTTCCCCG	$(TG)_{13}$	9	350–380	60	0.698	0.823	0.148	0.790	MF460398
Xtw11	F:TTACAAAACCTTGCTGTGCG R:GTTTTGATTACGACACCGCC	$(AC)_{30}$	10	250–290	60	0.600	0.882	0.001*	0.857	MF460399
Xtw17	F: CGCTTTTGAAATACCTTTTGA R: GCGTGTTCAAAGATGCTGAA	$(AG)_{15}$	11	180–220	60	0.382	0.843	0.030	0.810	MF460400
Xtw23	F: CGTCGTGACCTCTTGGCTAT R: GTAGTATGGCGGGAGGACAA	$(AT)_{12}$	13	240–310	60	0.682	0.919	0.000*	0.902	MF460401
Xtw47	F: TTGATGACAACAATAGGCG R: ACAAAAACCCTCCCCAAC	$(AT)_8$	7	188–208	60	0.467	0.763	0.000*	0.713	MF460402
Xtw48	F:TTCTACATGGCCGGATTG R:CTGATATCAACCCACGCA	$(AT)_{10}$	6	262–278	60	0.500	0.760	0.004	0.709	MF460403
Xtw51	F: GTTATCTGCATGCTTGGC R: GTGCAGAAAAAGTGCGAA	$(TG)_{15}$	16	236–276	60	0.591	0.873	0.000*	0.850	MF460404
Xth07	F: TGACTGGAACAAAAACGGAA R: ACGCACAACTGCACACATTT	$(AAT)_{13}$	10	150–186	60	0.870	0.868	0.137	0.843	MF460405
Xth09	F: GGACTGTAAGCGCTATGAGCA R: GAAGAAACGCTAATTTCTCTCCA	$(TAT)_{17}$	10	191–09	60	0.870	0.869	0.168	0.844	MF460406
Xth12	F: TCATCAAACGCCGTAAATCA R: GCTGAAAAGCGCTCCTAAAA	$(TGA)_{12}$	12	189–228	60	0.650	0.833	0.001*	0.859	MF460407
Xth17	F: CGCTCAAGACAATCTGTCCA R:TTTGACCTTAATTTGACGGC	$(TAA)_9$	13	199–238	60	0.711	0.868	0.030	0.843	MF460408
Xth19	F: CATGCAGTCGTTTAGGAGCA R: TGCCGTGTCAGTCAGTTCAT	$(AAT)_{13}$	17	169–220	60	0.783	0.936	0.003	0.921	MF460409
Xth22	F: CCTGTTCTTCCAGCAACGAT R: TGTTGCAGCAGTAGCATCAG	$(TAC)_9$	18	223–289	60	0.763	0.893	0.839	0.871	MF460410
Xth27	F: CTTCTGGGTGTGGAATTTGG R: TCTCCTACAGCGATGACGTG	$(TAT)_9$	10	226–253	60	0.609	0.860	0.004	0.834	MF460411
Xth29	F: CAAGTAGCAATGCAAATTCCC R: TATTCGTGCACGGTTATTGC	$(TTA)_9$	14	172–215	60	0.500	0.919	0.000*	0.899	MF460412
Xth31	F: TCATCAAACGCCGTAAATCA R: GATTTGCGCGTGAAATTAGG	$(TAC)_{28}$	15	166–217	60	0.894	0.928	0.147	0.912	MF460413

Table 1 (*contd*)

Locus	Primer sequence (5′–3′)	Repeat motif	No. of alleles	Size range (bp)	T_a (°C)	H_o	H_e	*P* value	PIC	GenBank accession number
Xth33	F: GGGAATCATCGCCAAAGTTA R: ACGTGAACGTTTCATTGCAG	$(ATT)_7$	11	172–202	60	0.439	0.872	0.000*	0.846	MF460414
Xth37	F:AATTCATGCAGCCGTTTAGG R:CAACTGTGAGATGACACGGG	$(AAT)_{11}$	11	248–281	60	0.200	0.857	0.000*	0.828	MF460415
Xth41	F: CATGCAGTCGTTTAGGAGCA R: TACATGATGCCCGTTTTTCA	$(AAT)_9$	12	198–234	60	0.717	0.895	0.020*	0.874	MF460416
Xth45	F: CATTTTGTTGCTGGACTGGA R: CGGTACGCACCATCTACTCA	$(TAT)_9$	16	166–211	60	0.783	0.919	0.004	0.902	MF460417
Xth47	F: ACCATCGTGGATTTTTACG R: CCTGTCTGTGGTGTAAGTG	$(ATG)_8$	13	243–288	60	0.761	0.880	0.480	0.858	MF460418
Xth54	F: CATATCATGCAGTCCATTG R: GGCCCTGTTCTTAAAATG	$(TCA)_{13}$	8	184–208	60	0.745	0.826	0.451	0.793	MF460420

T_a, annealing temperature; H_o, observed heterozygosity; H_e, expected heterozygosity; PIC, polymorphism information content. *Significant deviations from HWE after Bonferroni correction ($P < 0.05/23$).

microsatellite development in nonmodel species without genomic information.

Twenty-three microsatellite loci with high polymorphism were screened successfully (table 1). The number of alleles per locus ranged from 6 to 18. The observed and expected heterozygosities varied from 0.200 to 0.894 and 0.760 to 0.936, respectively. Nine loci deviated significantly from HWE after Bonferroni correction, which might be expected because of the presence of null alleles. In marine molluscs, many microsatellites suffer from surprisingly high levels of null alleles (Reece *et al.* 2004; Yu and Li 2008; Ni *et al.* 2011; Gao *et al.* 2016), which usually cause deviation from HWE. No linkage disequilibrium was detected among the loci ($P > 0.01$). All the values of PIC were >0.5, suggesting that these microsatellite loci were highly polymorphic. These microsatellite makers will be useful for assessing genetic diversity and population structure of *C. rustica* and developing the effective conservation and management strategies.

Acknowledgements

This study was supported by the grants from Fundamental Research Funds for the Central Universities (201762014) and Industrial Development Project of Qingdao City (17-3-3-64-nsh).

References

Baird N. A., Etter P. D., Atwood T. S., Currey M. C., Shiver A. L., Lewis Z. A. and Johnson E. A. 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Berman M., Austin C. M. and Miller A. D. 2014 Characterisation of the complete mitochondrial genome and 13 microsatellite loci through next-generation sequencing for the New Caledonian spider-ant *Leptomyrmex pallens*. *Mol. Biol. Rep.* **41**, 1179–1187.

Dieringer D. and Schlötterer C. 2003 Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **31**, 167–169.

Gao X., Zheng X., Bo Q. and Li Q. 2016 Population genetics of the common long-armed octopus *Octopus minor* (Sasaki, 1920) (Cephalopoda: Octopoda) in Chinese waters based on microsatellite analysis. *Biochem. Syst. Ecol.* **66**, 129–136.

Kalinowski S. T., Taper M. L. and Marshall T. C. 2007 Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**, 1099–1106.

Morgante M., Hanafey M. and Powell W. 2002 Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* **30**, 194–200.

Ni L., Li Q. and Kong L. 2011 Microsatellites reveal fine-scale genetic structure of the Chinese surf clam. *Mactra chinensis* (Mollusca, Bivalvia, Mactridae) in Northern China. *Mar. Ecol.* **32**, 488–497.

Reece K. S., Ribeiro W. L., Gaffney P. M., Carnegie R. B. and Allen S. K. Jr. 2004 Microsatellite marker development and

analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *J. Hered.* **95**, 346–352.

Rousset F. 2008 Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Res.* **8**, 103–106.

Thiel T., Michalek W., Varshney R. K. and Graner A. 2003 Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422.

Wang Y. and Wei Y. 1994 Ecological investigation of Trochidae (Mollusca) in the intertidal zone of Zhoushan islands, Zhejiang. *J. Zhejiang Coll. Fisheries* **13**, 38–44 (in Chinese).

Yu H. and Li Q. 2008 Exploiting EST databases for the development and characterization of EST-SSRs in the Pacific oyster (*Crassostrea gigas*). *J. Hered.* **99**, 208–214.

Corresponding editor: INDRAJIT NANDA