ONLINE RESOURCES

Isolation and characterization of twenty-nine novel EST-SSR markers in *Siniperca undulata*

CHUN-MEI QU, XU-FANG LIANG*, WEI HUANG, CHENG ZHAO and LIANG CAO

College of Fisheries, Huazhong Agricultural University, Wuhan 430070, People's Republic of China

[Qu C.-M., Liang X.-F., Huang W., Zhao C. and Cao L. 2013 Isolation and characterization of twenty-nine novel EST-SSR markers in *Siniperca undulata*. *J. Genet*. **92**, e116–e120. Online only: http://www.ias.ac.in/jgenet/OnlineResources/92/e116.pdf]

Introduction

Siniperca undulata Fang and Chong is a member of the genus *Siniperca* whose population genetic structure remains unknown. In this study, we report the development of the first set of microsatellite markers for *S. undulata* from the transcriptome of F_1 hybrids between *Siniperca chuatsi* (♀) and *Siniperca scherzeri* (♂). The level of genetic variability was assessed for 30 *S. undulata* individuals. Twenty-nine loci were polymorphic, with the number of alleles per locus ranging from 3 to 13. The observed and expected heterozygosity ranged from 0.0667 to 1.0000 and from 0.2130 to 0.9006, respectively. Two loci deviated from Hardy–Weinberg equilibrium (HWE) and two loci in significant linkage disequilibrium (LD) were detected. These novel markers will facilitate further studies on population genetic structure, genetic diversity and conservation genetics in this species.

S. undulata is mainly distributed in China (Zhou *et al*. 1988) and shows relatively close relationship to *S. scherzeri*. *S. undulata* is considered Near Threatened (Zhao 2011) because it has been impacted by overfishing and pollution, and its population decline is suspected to be close to 30% over the past 10 years. There is thus a significant need to protect this species. So far, few researches were reported about *S. undulata* and the population genetic structure remains unknown. However, understanding the population structure is important to find appropriate conservation strategies of biodiversity (Cegelski *et al*. 2003).

Microsatellites or simple sequence repeats (SSRs) have become a useful marker system in population genetics analysis due to their codominant nature and high allelic polymorphism (Walter and Epperson 2001). Recent advances in next-generation sequencing technologies allow large scale generation of ESTs efficiently and cost-effectively (Metzker 2010). Microsatellite markers have been isolated for many important sinipercine fish species such as *S. chuatsi* (Zhang *et al*. 2006; Kuang *et al*. 2009; Liu *et al*. 2011; Huang *et al*. 2012a; Qu *et al*. 2012, 2013) and *S. scherzeri* (Huang *et al*. 2012a, b; Qu *et al*. 2012, 2013; Yang *et al*. 2012). Nevertheless, to date no polymorphic microsatellites were developed for *S. undulata* because of the lack of genomic resources such as genome or transcriptome sequences. Thus, the development of microsatellite markers are urgently needed to allow genetic diversity studies permitting the development of conservation strategies for this species. Here, we develop the first set of EST-SSR markers for *S. undulata*.

Materials and methods

BatchPrimer3 ver. 1.0 software was used to mine EST-SSR markers from the transcriptome sequences which were obtained in our previous study (Qu *et al*. 2012) with the default parameters (You *et al*. 2008). A total of 22,418 SSR loci were identified and 73 were selected for microsatellite marker optimization in this study. Primers for these loci were designed using NCBI/Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Genomic DNA was extracted from fin clips using the TIANamp Genomic DNA Kit (Tiangen, Beijing, China) following the manufacturer's instructions. The polymorphism of microsatellite loci was evaluated on 30 individuals from a wild *S. undulata* population. Polymerase chain reaction (PCR) conditions were optimized for each pair of primers. PCRs were performed in 25 μL reaction volumes containing 2.5 μL of 10× PCR buffer, 1.0–3.0 mM $MgCl_2$, 50 μM dNTPs, 0.4 μM of each primer, 1 U *Taq* polymerase (Takara, Tokyo, Japan) and 50 ng genomic DNA. PCR conditions

*For correspondence. E-mail: xfliang@mail.hzau.edu.cn.

Keywords. microsatellite; EST-SSR; genetic diversity; *Siniperca undulata*.

Table 1. Characterization of 29 polymorphic EST-SSR markers in *S. undulata*.

Locus	Accession number	Repeat motif	Primer sequence (5′–3′)	Size range (bp)	T_a (°C)	N_a	H_O	H_E	PIC	*P* value	N_s
SU267	JX443420	$(AC)_{11}$	F: AAGGCAGGATGCACTGAGTC R: ACTTAAGACCATACACCTGGAGA	177–249	55.0	9	0.9667	0.8633	0.8306	0.3787	30
SU268	JX443421	$(TG)_{11}$	F: TCACCAGTTCACCCAGTGGA R: GCGATACCGCTCGAGGAAAA	158–203	55.0	7	0.7333	0.7379	0.6881	0.6878	30
SU269	JX443422	$(TG)_{11}$	F: TACCACGGGGTACTGGACTG R: GCAATAAAAACCTACCCCATCCG	140–193	55.0	8	0.9000	0.8644	0.8319	0.4307	29
SU270	JX443423	$(AC)_{11}$	F: GTGGCCTGGTTTGTGGTTCC R: AGCCGCATGAATCCAGTCCT	117–181	55.0	8	1.0000	0.7040	0.6502	0.0000*	30
SU271	JX443424	$(CA)_{11}$	F: TGACAACCAATTTACCCATTGCT R: TCAAGTTCTGGACTTACACCAGC	156–191	55.0	4	0.9000	0.6503	0.5792	0.0060	28
SU272	JX443425	$(TC)_{11}$	F: GCATTGGGAAGTGGCACAGT R: AGGCATGACGTGCATCTGAA	112–139	55.0	3	0.1000	0.2672	0.2415	0.9999	30
SU273	JX443426	$(TG)_{11}$	F: TGGTATTGGCAAAGCAATCCCT R: CCGAGACATCCATTGAGGCA	139–158	55.0	5	0.8000	0.7169	0.6600	0.3928	30
SU275	JX443427	$(CA)_7CG(CA)_{11}$	F: ACCGAACAAGAAAAGACTCCTGG R: GCTGGCCTCTCCTCTTGGA	127–196	55.0	8	0.9333	0.8429	0.8063	0.3511	30
SU279	JX443428	$(GAT)_7$	F: CACTTGGGGCTCTACATTGC R: TACAGAGGCGCAGTCACTCA	90–144	55.0	13	0.9333	0.9006	0.8756	0.9729	29
SU281	JX443429	$(TCC)_7$	F: CAACTGCAGAATGGTTCGCC R: GGACGGACAGCAAGGCTAC	87–99	55.0	3	0.3333	0.4881	0.4122	0.9955	29
SU285	JX443430	$(GAT)_7$	F: CGGAGGCACGACATGAAAGT R: AGATATCCTGCACCACCCCC	109–141	55.0	7	0.6000	0.6938	0.6461	0.8714	30
SU287	JX443431	$(TTA)_7$	F: TTGCCCAAGGTAGGGTCTGA R: TGTGCCTTCTCTTTACCCAAGG	100–127	47.2	7	0.6000	0.7056	0.6573	0.9988	27
SU288	JX443432	$(TCA)_7$	F: ACAGACCTCTCGGTTTTGTCA R: ATGCTGGAGGTGCATGTGTTT	151–220	55.0	6	0.7000	0.6938	0.6353	0.9631	30
SU289	JX443433	$(CAG)_7$	F: TCAGAGACTCAGAGAGAAGAAGTC R: GTGGCCACAGTGGGGGA	218–238	55.0	3	0.4000	0.5859	0.5065	0.9867	30
SU290	JX443434	$(TGG)_7$	F: TGTACTCGTCTTCCACCCCC R: AGCCGCTGGTCACGGA	154–195	55.0	3	0.2333	0.2130	0.1935	0.6663	30
SU294	JX443435	$(TCA)_7$	F: AAGCTGCTGAACGTCGCCTC R: GCAGGCTCAGACCTCCTCCAT	115–188	55.0	7	0.6000	0.6927	0.6439	0.9983	29
SU298	JX443436	$(CCA)_7$	F: TAGCCATTATTTACTCGGAG R: GCTTGATCTGTAGGGAGTGC	109–137	55.0	6	0.4667	0.8056	0.7616	1.0000	30
SU305	JX443437	$(GAA)_7$	F: ACTGGGGTCAGCTTTTTCAC R: TCATCAATAAACTGCGGT	122–146	55.0	4	0.0667	0.5565	0.4901	0.0000*	29
SU306	JX443438	$(TTC)_7$	F: AGCTTCTTCTCTTTGTCCTG R: GAGTTTGGAGACGATGAC	268–297	55.0	4	0.3333	0.6768	0.6127	1.0000	29
SU308	JX443439	$(ATT)_7$	F: AAGCTCTGCTGCTTGTCTATGA R: CACCATCTCGTTTTCAATGGGG	206–234	55.0	3	0.3667	0.5237	0.4244	0.9324	30
SU312	JX443440	$(GAT)_7$	F: ACTAACGCAGCTTCGGTCTG R: TTGTTTTGTACTCTTGGCTTCCT	64–84	55.0	8	0.7000	0.8418	0.8058	0.9808	29

Table 1. (*contd*).

Locus	Accession number	Repeat motif	Primer sequence (5′–3′)	Size range (bp)	T_a (°C)	N_a	H_O	H_E	PIC	*P* value	N_s
SU313	JX443441	$(GAG)_7$	F: CCGAAGTGAAGATGCAGAGGG R: ATCTCGGCTGGACCGTGAA	202–239	55.0	4	0.5000	0.7079	0.6483	0.9951	30
SU315	JX443442	$(TTC)_7$	F: GATGTTTCCCGCCTCGATCC R: GCATTTTTCCCAGCAGCCTC	193–253	55.0	5	0.5333	0.7497	0.6910	0.9991	28
SU318	JX443443	$(AAT)_7(TAA)_6$	F: ATAAGCCTATACAGCCACT R: TGTTGTTCCATCTCCTCA	168–226	55.0	7	0.7333	0.7960	0.7558	0.5491	30
SU320	JX443444	$(GAG)_7$	F: TGCTGGTGAAACTCTCCCCG R: GGCCTTGCAAAAGAAAGTGCT	180–219	55.0	4	0.4000	0.6480	0.5823	0.9975	30
SU322	JX443445	$(CAT)_7$	F: TCTACCACAGCTGCCAGTCA R: GCTGACTCACTCCTCGGGAT	188–217	53.5	4	0.4333	0.5689	0.4992	0.9345	30
SU333	JX443446	$(ATGT)_5$	F: CTGGCCTGCCAGAACAAAAT R: TGGCCTCTTCAGACACATTGC	206–247	55.0	4	0.3333	0.5847	0.5058	0.9901	30
SU334	JX443447	$(GA)_{10}$	F: AACGAGGCATCTTCGTCCCT R: CGGGGATACACTCCGGTACA	148–164	55.0	3	0.4333	0.3621	0.3188	0.1896	30
SU337	JX443448	$(TCA)_7$	F: GAGGAAAAGGCTTGTGGGCT R: GTGCAGATGTGTCTTGTTTTGGA	150–186	55.0	6	0.5000	0.8062	0.7609	0.9943	29

T_a, annealing temperature; N_a, number of alleles; H_O/H_E, observed and expected heterozygosity; PIC, polymorphic information content; N_s, number of successfully genotyped samples.
*Significant deviation from HWE after Bonferroni correction ($P < 0.0017$).

were as follows: initial denaturation at 94°C for 3 min followed by 35 cycles at 94°C for 30 s, the optimized annealing temperature (table 1) for 30 s, 72°C for 30 s, and then a final extension step at 72°C for 10 min. PCR products were separated on a 8% nondenaturing polyacrylamide gel electrophoresis and visualized by silver staining. A denatured pBR322 DNA/*Msp*I molecular weight marker (Tiangen) was used as a size standard to identify alleles.

Genetic diversity parameters of polymorphic loci were calculated using PopGene ver. 1.32 (Yeh and Boyle 1997), such as the number of alleles (N_a), the observed (H_O) and expected heterozygosity (H_E). We tested for deviations from HWE (i.e. heterozygote deficiency and heterozygote excess) and LD using the online version of Genepop (http://genepop.curtin.edu.au/) (Raymond and Rousset 1995). Markov chain parameters were left at the default settings. All results were adjusted for multiple simultaneous comparisons using a sequential Bonferroni correction (Rice 1989). We also calculated the polymorphic information content (PIC) using the formula:

$$\mathrm{PIC} = 1 - \left(\sum_{i=1}^{n} q_i^2\right) - \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2q_i^2 q_j^2\right),$$

where n is the number of alleles, and q_i and q_j is the ith and jth allele frequency, respectively (Botstein *et al*. 1980). Genotyping errors due to null alleles, stutter bands, or allele dropout were checked by Micro-Checker ver. 2.2.3 software (Van-Oosterhout *et al*. 2004).

Results and discussion

As shown in table 1, 29 of the 73 tested loci were polymorphic. For the polymorphic loci, the number of alleles observed ranged from 3 to 13, with observed heterozygosity (H_O) ranging from 0.0667 to 1.0000 (average of 0.5701), and expected heterozygosity (H_E) from 0.2130 to 0.9006 (average of 0.6637), respectively. Although the variability observed in these EST-SSRs is possibly underestimated due to small sample size (30 individuals), it still showed moderate to high levels of polymorphism. A total of 27 loci conformed to HWE, and only two (SU270 and SU305) deviated from HWE after Bonferroni correction (adjusted $P = 0.0017$), which may be due to the small sample size ($n = 30$) or nonrandom mating. Two (SU270 and SU305) loci in significant LD were detected across all loci following Bonferroni correction (adjusted $P = 0.0001$). No evidence for allelic dropout and scoring error due to stuttering was found in these loci. Null alleles may be present at nine loci.

PIC is an important parameter used in the practicability test of the designed SSR markers. In the 30 examined *S. undulata* samples, the PIC value of the 29 newly developed EST-SSR markers ranged in a quite broad region, from 0.1935 to 0.8756, and the average was 0.6109 (table 1). These results indicate that the developed EST-SSR markers

Table 2. Annotation of gene-associated SSRs by basic local alignment search tool (BLAST) analysis.

Locus	BLASTx similarity match	*e*-Value	Species	Accession number
SU267	GATS-like protein 3-like	8E-46	*Maylandia zebra*	XM_004566701.1|
SU268	3-Hydroxyacyl-CoA dehydratase-like	2E-71	*M. zebra*	XM_004573284.1
SU270	T calcium channel alpha 1G subunit varian 88	1E-11	*Dicentrarchus labrax*	FQ310507.3
SU271	Integrin alpha-7	4E-12	*D. labrax*	FQ310506.3
SU272	Phospholipase C delta 3A	1E-124	*Paralichthys olivaceus*	EU433322.1
SU273	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit C-like	0	*M. zebra*	XM_004560422.1
SU275	B-cell CLL/lymphoma 9 protein-like, transcript variant X2	0	*M. zebra*	XM_004557883.1
SU279	Apoptosis-stimulating of p53 protein 2-like, transcript variant X2	3E-91	*M. zebra*	XM_004551716.1
SU281	Short stature homeobox protein 2-like, transcript variant X2	6E-70	*M. zebra*	XM_004543712.1
SU285	Hypothetical protein LOC100698742 (LOC100698742)	5E-60	*Oreochromis niloticus*	XM_003443043.1
SU288	Transcription factor Sp3-like, transcript variant X2	1E-18	*M. zebra*	XM_004540307.1
SU289	DNA (cytosine-5)-methyltransferase 3A-like, transcript variant X2	4E-91	*M. zebra*	XM_004540232.1
SU290	Transcriptional repressor protein YY1-like, mrna	1E-84	*Takifugu rubripes*	XM_003971528.1
SU294	ETS domain-containing protein Elk-1-like, transcript variant X5	2E-45	*M. zebra*	XR_191643.1
SU305	Forkhead box protein L1	0	*D. labrax*	FQ310507.3
SU306	ESF1 homolog, transcript variant X2	3E-166	*M. zebra*	XM_004538635.1
SU312	Uncharacterized LOC101469285, transcript variant X2	1E-34	*M. zebra*	XM_004552712.1
SU313	Homeobox protein unc-4 homolog	5E-116	*M. zebra*	XM_004569064.1
SU315	U4/U6 small nuclear ribonucleoprotein Prp4-like, transcript variant X3	9E-55	*M. zebra*	XM_004556182.1
SU318	Protein slowmo homolog 2-like	1E-48	*M. zebra*	XM_004554056.1
SU320	Insulinoma-associated protein 1-like	0	*M. zebra*	XM_004566541.1
SU322	Adenomatous polyposis coli homolog	0	*O. niloticus*	XM_003453615.1
SU333	Receptor expression-enhancing protein 1-like, transcript variant X3,	2E-60	*M. zebra*	XM_004540587.1
SU334	Atrial natriuretic peptide receptor 1-like	3E-57	*M. zebra*	XM_004575232.1
SU337	Unconventional myosin-X-like	5E-30	*M. zebra*	XM_004569454.1

are of good practicability and the 30 examined *S. undulata* samples are of good genetic diversity.

To determine the function of genes associated with the SSR markers, BLASTx (basic local alignment search tool x, http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) searches were conducted for all unigenes with polymorphic SSRs using cut-off *E* value < 1.00E-10 (Sha *et al*. 2011). Of 29 SSR-containing unigenes, 25 had significant homology to known genes. The results of the BLAST are shown in table 2.

Here, we tested 73 microsatellite loci for *S. undulata* from the transcriptome of F_1 hybrids between *S. chuatsi* and *S. scherzeri*, of which 59 (80.8%) were successfully amplified and 29 (39.7%) were polymorphic. The percentage of polymorphic loci was moderate, indicating the potential utility of transcriptome sequences for the identification and characterization of large numbers of gene-based SSR loci across species for which limited marker resources were available. These microsatellite loci were the first set of EST-SSRs in *S. undulata* and would be valuable for studies of genetic diversity, population structure, conservation genetics, selective breeding programmes, parentage and genome mapping.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (31172420), the National Basic Research Program of China (2009CB118702) and the Fundamental Research Funds for the Central Universities (2010PY010, 2011PY030).

References

Botstein D., White R. L., Skolnick M. and Davis R. W. 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331.

Cegelski C. C., Waits L. P. and Anderson N. J. 2003 Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Mol. Ecol.* **12**, 2907–2918.

Huang W., Liang X. F., Qu C. M., Cao J. J., Zhao C. and Cao L. 2012a Development and characterization of novel polymorphic microsatellite loci in *Siniperca scherzeri* Steindachner and *Siniperca chuatsi* (Basilewsky). *Mol. Biol. Rep.* **40**, 751–756.

Huang W., Liang X. F., Qu C. M., Li J. and Cao L. 2012b Isolation and characterization of twenty-five polymorphic microsatellite markers in *Siniperca scherzeri* Steindachner and cross-species amplification. *J. Genet.* **91**, e113–e117.

Kuang G. Q., Lu S. Q., Zheng S. M. and Wu Q. 2009 Isolation and evaluation of 18 microsatellite loci in *Siniperca chuatsi* (Basilewsky). *Mol. Ecol. Res.* **9**, 1473–1475.

Liu X. L., Luo W., Zeng C., Wang W. M. and Gao Z. X. 2011 Isolation of new 40 microsatellite markers in mandarin fish (*Siniperca chuatsi*). *Int. J. Mol. Sci.* **12**, 4180–4189.

Metzker M. L. 2010 Sequencing technologies-the next generation. *Nat. Rev. Genet.* **11**, 31–46.

Qu C. M., Liang X. F., Huang W. and Cao L. 2012 Isolation and characterization of 46 novel polymorphic EST-simple sequence repeats (SSR) markers in two sinipercine fishes (*Siniperca*) and cross-species amplification. *Int. J. Mol. Sci.* **13**, 9534–9544.

Qu C. M., Liang X. F., Huang W., Zhao C., Cao L., Yang M. and Tian C. X. 2013 Development and characterization of twenty-nine novel polymorphic microsatellite loci in the mandarin fish *Siniperca chuatsi*. *J. Genet.* **92**, e19–e23.

Raymond M. and Rousset F. 1995 GENEPOP (version 1.2): Population genetic software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249.

Rice W. R. 1989 Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Sha Z. X., Luo X. H., Liao X. L., Wang S. L., Wang Q. L. and Chen S. L. 2011 Development and characterization of 60 novel EST-SSR markers in half-smooth tongue sole *Cynoglossus semilaevis*. *J. Fish Biol.* **78**, 322–331.

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.

Walter R. and Epperson B. K. 2001 Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. *Mol. Ecol.* **10**, 103–111.

Yang M., Liang X. F., Tian C. X., Gul Y., Dou Y. Q., Cao L. and Yu R. 2012 Isolation and characterization of fifteen novel microsatellite loci in golden mandarin fish (*Siniperca scherzeri*) Steindachne. *Conserv. Genet. Resour.* **4**, 599–601.

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

You F. M., Huo N., Gu Y. Q., Luo M. C., Ma Y., Hane D. *et al*. 2008 BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**, 253.

Zhang B., Li Z. J., Tong J. G. and Liao X. L. 2006 Isolation and characterization of 18 polymorphic microsatellite markers in Chinese mandarin fish *Siniperca chuatsi* (Basilewsky). *Mol. Ecol. Notes* **6**, 1216–1218.

Zhao H. 2011 *Siniperca undulata*. In *IUCN 2012. IUCN red list of threatened species*. Version 2012.2 (www.iucnredlist.org).

Zhou C., Yang Q. and Cai D. L. 1988 On the classification and distribution of the Sinipercinae fishes (family Serranidae). *Zool. Res.* **9**, 113–125 (in Chinese).

Received 6 March 2013, in revised form 2 July 2013; accepted 8 July 2013
Published on the Web: 11 December 2013