



Development of polymorphic microsatellites for *Sillago sihama* based on next-generation sequencing and transferability to *Sillago japonica*

R.X. Wu, H.R. Zhang, S.F. Niu, Y. Zhai and X.F. Liu

College of Fisheries, Guangdong Ocean University, Zhanjiang, China

Corresponding author: S.F. Niu

E-mail: wolf0487@126.com

Genet. Mol. Res. 15 (4): gmr15049046

Received July 29, 2016

Accepted September 19, 2016

Published November 21, 2016

DOI <http://dx.doi.org/10.4238/gmr15049046>

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. *Sillago sihama* (Forsskål, 1775), a commercially important marine fishery species in the Indo-West Pacific, is being developed as a target species for aquaculture and stock enhancement in China. However, due to the limitations of traditional isolation methods, the available microsatellite loci, or simple sequence repeats (SSRs), of *S. sihama* lack diversity. We used a stepwise approach including Illumina sequencing, primer screening, and SSR marker validation to develop diverse SSRs for *S. sihama*. A total of 853.48 Mb clean sequences were assembled with high coverage and sequencing depth, and 27,288 potential SSRs were identified. A set of 18 novel SSR markers with four type motifs including 11 di-, 1 tri-, 5 tetra-, and 1 hexanucleotide repeats were successfully isolated. The ranges of number of alleles per locus and observed and expected heterozygosities were 5-24, 0.226-0.968, and 0.319-0.950, respectively. The diversity parameters exhibited high

levels of polymorphism in these 18 loci. Three loci with the presence of both null alleles and inbreeding showed significant deviation from Hardy-Weinberg equilibrium after Bonferroni correction. Moreover, 13 loci developed in *S. sihama* showed high transferability to the closely related species *Sillago japonica*. The polymorphic SSR markers developed in this study may serve as valuable tools for further basic and applied research on the genetic resources of *S. sihama* as well as *S. japonica*. Our results indicate that this approach, based on next-generation sequencing technology, is convenient, cost-effective, and suitable for SSR marker isolation in other sillaginid fishes.

Key words: *Sillago sihama*; *Sillago japonica*; Microsatellite loci; Next-generation sequencing; Loci transferability

INTRODUCTION

Due to their ubiquitous occurrence and high allelic polymorphism, microsatellite loci, or simple sequence repeats (SSRs), have been widely applied in various research fields, especially in population genetics, conservation biology, molecular ecology, and genetic breeding (Ma et al., 2012; Zhang et al., 2016). However, traditional approaches for SSR isolation are laborious, time-consuming, and costly (Abdelkrim et al., 2009). Recently, with the rapid development of next-generation sequencing (NGS) technologies and their novel biological applications (Grover and Sharma, 2016), it is becoming more and more convenient to generate a large number of SSR reads from the genome in a short time and at a relatively low cost (Duan et al., 2014; Yang et al., 2015). Furthermore, genome *de novo* sequencing, genome re-sequencing, and RNA sequencing can also be realized in model or non-model species based on NGS technologies (Schuster 2008; Yang et al., 2015). Indeed, NGS technologies have been successfully applied to develop SSR markers for a large variety of organisms without reference sequences (Addamo et al., 2015; Villanova et al., 2015), and even for species such as birds, which have few SSRs (Page et al., 2014), and the extinct moa, which has a small quantity of ancient DNA (Allentoft et al., 2009).

The family Sillaginidae (Teleostei, Percoidei) contains four genera in approximately 33 species (Shao, 2016). They are small to moderate in size and typically inhabit the inshore coastal waters with sandy substrates or estuaries (Gao et al., 2011). As a representative species of this family, *Sillago sihama* (Forsskål, 1775) is the most widespread and abundant member of the sillaginids and is distributed in the Indo-West Pacific region from Knysna (South Africa) to Japan (Liu et al., 2012). *S. sihama* can be found in all Chinese coastal waters, supporting a commercially important marine fishery. Currently, the supply of this fish mainly relies on wild capture (Guo et al., 2012). However, *S. sihama* has been overexploited in the East and South China Seas during the last 20 years (Lu et al., 2008; Liu et al., 2010). This requires us to pay close attention to the genetic diversity conservation of wild *S. sihama*. Furthermore, due to the successful conduction of artificial propagation in the South China Sea (Huang et al., 2013), *S. sihama* is being developed as a target species for aquaculture and stock enhancement. Thus, escapes or releases of hatchery-produced *S. sihama* may pose a potential genetic risk to local populations. Accordingly, it is necessary to develop polymorphic SSR markers in order to provide an essential basis for future genetic resource study and aquaculture practice in *S. sihama*.

The SSRs of *S. sihama* and two other *Sillago* species (*Sillago japonica* Temminck & Schlegel, 1843, and *Sillago parvisquamis* Gill, 1861) have been isolated using different molecular techniques (Guo et al., 2012; Ueno et al., 2013; Umino et al., 2013; Wang et al., 2014), but none of them were developed from NGS technologies. The 66 SSRs developed in *S. sihama* using the FIASCO method, however, were all dinucleotide repeats, and 42 of them showed low polymorphism, with no more than six alleles (Guo et al., 2012). Moreover, traditional methods based on the enriched genomic library usually obtain a specific microsatellite type. For example, only dinucleotide repeat microsatellites of both *S. sihama* and *S. japonica* were isolated in previous reports (Guo et al., 2012; Ueno et al., 2013; Wang et al., 2014). This may limit both the diversity of SSR markers and the number of polymorphic SSRs obtained from the genome (Restrepo et al., 2015). Given these deficiencies, we utilized the high-throughput Illumina sequencing platform to develop a set of diverse SSR markers for *S. sihama*.

The aim of this study was to i) identify genomic SSRs and characterize 18 polymorphic SSR markers in *S. sihama*, ii) test the transferability of these new markers to the closely related species *S. japonica*, and iii) establish the first complete SSR marker isolation procedure in *S. sihama* using NGS technology. These results will significantly contribute to understanding the population genetic resources and conservation of *S. sihama* as well as *S. japonica*, and could also provide an alternative approach to the SSR marker isolation of other sillaginid fishes.

MATERIAL AND METHODS

Sample collection and DNA extraction

Fresh *S. sihama* (32 individuals) and *S. japonica* (24 individuals) were sampled from the east and west coasts of Leizhou Peninsula, Guangdong Province, the South China Sea, respectively. Samples were preserved in 95% alcohol, and total genomic DNA was extracted from muscle as described by Sambrook and Russell (1989).

Illumina paired-end sequencing, data filtering and genome assembly

One specimen of *S. sihama* was subjected to high-throughput sequencing using an Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA) at BGI Genomics Co., Ltd. (Shenzhen, China), and a 100-bp read length paired-end sequencing library was constructed with insert sizes of about 500 bp. Illumina sequencing generated 1632 Mb of raw genomic data with 9.80X sequencing depth. Through a stringent filtering process (Li et al., 2010), which included removing reads containing adapter sequences, low quality reads (quality score of $Q \leq 5$), reads containing poly-N, and duplicate reads (Duan et al., 2014), about 8% of raw reads were cleaned. The remaining 1479.99 Mb reads were high quality with 9.06X sequencing depth. Due to the low frequency (< 3) K-mer value ($K = 17$ bp) (Li et al., 2010), a small dataset of error sequencing reads was then removed from the raw reads (about 2.11%). Genome *de novo* assembly was performed using the SOAPdenovo software package with default settings (Duan et al., 2014). After contig assembly (contig longer than 100 bp were used), scaffolding, and gap closure (Li et al., 2010), the resulting 853.48 Mb assembled genome datasets were used for SSR discovery.

SSR identification and primer design

The potential SSRs were searched using the Tandem Repeats Finder tool (Benson, 1999) (<http://tandem.bu.edu/trf/trf.html>) according to the following criteria: dinucleotide motifs with ≥ 8 repeats, trinucleotide motifs with ≥ 6 repeats, tetranucleotide motifs with ≥ 4 repeats, and pentanucleotide and hexanucleotide motifs with ≥ 3 repeats. To obtain high-quality polymerase chain reaction (PCR) primers, we used stringent criteria for primer design: i) amplification products between 180 and 300 bp, ii) primer lengths of 17-23 bp with a maximum 3 bp length difference between paired primers, iii) GC content of 45-60%, and iv) melting temperature of 50°-65°C. Primers were designed by Primer Premier 5.0 (Premier Biosoft International, Palo Alto, CA, USA). The primer pair with the highest Premier 5.0 assigned score was chosen for synthesis (Sangon Biotech Co., Ltd., Shanghai, China).

Primer polymorphism test

A total of 133 primer pairs containing di- to hexanucleotide repeats were randomly selected. Polymorphism detection was conducted in six *S. sihama* individuals. PCR was carried out in 15 μL volumes containing 1X PCR buffer [100 mM Tris-HCl pH 8.8, 25°C, 500 mM KCl, 0.8% Nonidet P40 (v/v), 15 mM MgCl_2], 1 μL extracted DNA (10-50 ng), 0.3 μL each primer (10 μM), 0.3 μL dNTP mixture (10 mM), and 1.7 U *Taq* DNA polymerase (Transgen Biotech Co., Ltd., Beijing, China). The PCR amplifications were performed on a Veriti™ 96-Well Thermal Cycler (Applied Biosystems, USA) under the following conditions: an initial denaturation at 95°C for 4 min, a touchdown PCR protocol for 25 cycles at 95°C for 30 s, 62°-52°C for 30 s (decreasing 1°C per cycle for the first 10 cycles), 72°C for 30 s, and a final extension at 72°C for 10 min. PCR products were separated on 8% non-denaturing polyacrylamide gel and visualized by 0.1% silver nitrate staining. The allele size was identified roughly according to the pBR322 Marker/*MspI* marker (Tiangen, Beijing, China).

PCR amplification and genotyping

According to the M13 labeling method (Schuelke, 2000), an M13 tail (5'-AGGGTTTTCCAGTCACG-3' or 5'-GAGCGGATAACAATTTTCACAC-3') was added to the 5' end of each forward primer of selected polymorphic SSRs (Table 1), while the M13 universal primer (the sequence was same as for the M13 tail) was labeled with FAM or HEX (Generay Biotech Co., Ltd., Shanghai, China) at its 5' end. The optimal annealing temperature was identified for each pair of primers via gradient PCR (51°, 53°, 55°, 57°, 59°, and 61°C) using the three-primer PCR amplification protocol described by Boutin-Ganache et al. (2001) and Schuelke (2000). PCR amplifications were performed in a total volume of 15 μL , which included 1X PCR buffer [100 mM Tris-HCl, pH 8.8, 25°C, 500 mM KCl, 0.8% Nonidet P40 (v/v), 15 mM MgCl_2], 1 μL extracted DNA (10-50 ng), 0.3 μL forward primer (10 μM) with M13 tail, 0.1 μL labeled M13 universal primer (10 μM), 0.3 μL reverse primer (10 μM), 0.3 μL dNTP mixture (10 mM), and 1.7 U *Taq* DNA polymerase (Transgen Biotech Co., Ltd., Beijing, China). The PCR programs were conducted as follows: an initial denaturation at 95°C for 5 min, followed by 25 cycles at 95°C for 30 s, 30 s at the locus-specific annealing temperature (51°-61°C), 30 s at 72°C, and a final extension for 10 min at 72°C. All sets of PCR included a negative control reaction tube in which all reagents were included except the

template DNA. The successful PCR products were sent to Shanghai Generay Biotech Co., Ltd. (Shanghai, China) for genotyping by an Applied Biosystems 3730 DNA Analyzer and GeneMapper[®] software version 4.0.

Table 1. Eighteen microsatellite primers developed in *Sillago sihama* and 13 primer pairs from successful cross-amplification of *Sillago japonica*.

Locus	Primer sequences(5'-3')	Repeat motif	Allele size (bp)		Ta (°C)	GenBank accession No.
			<i>S. sihama</i>	<i>S. japonica</i>		
SS2-1	F: FAM-TGTCCACAGTGAAGCAGA R: TATGAGCCAAATCAGAAAGGTC	(CA) ₁₃	254-348	–	57	KX074059
SS2-2	F: FAM-GAGGAAGAAAGAAAGGGAGT R: TGC GTTATCTGGGTGT	(AG) ₁₁	206-242	188-214	53	KX074060
SS2-3	F: FAM-AGGTGTCGTCTGGCTTCA R: TGGGGCTGGTTCGGGTT	(CT) ₁₂	206-292	229-257	57	KX074061
SS2-4	F: HEX-CAATGTGCTCGCAAAGACG R: CGCCGCTATTCATAAACCC	(CA) ₁₁	203-247	200-230	58	KX074062
SS2-6	F: FAM-GTGACATAAATGCTCATCTGC R: TGT TTTGTGTGCGACTTG	(TG) ₁₁	212-274	189-205	53	KX074063
SS2-10	F: HEX-ATGCTGGCAGAGGCTTTA R: AGGGAACGGCAGATTGA	(TG) ₁₄	128-182	152-198	59	KX074064
SS2-11	F: FAM-AGCGTTGTCATCACCTGC R: ATCCCTGAGCCCATCT	(AC) ₄	257-313	253-297	53	KX074065
SS2-12	F: HEX-TTCAGCCGCTTATCAATC R: ACTCCAGGACCTTTACC	(GT) ₁₅	164-202	158-204	59	KX074066
SS2-13	F: FAM-AAACAGCGAAGGCACATA R: GACATTTACGGAGCATC	(TC) ₁₃	184-232	174-216	51	KX074067
SS2-14	F: FAM-CTGAGGGCAGCACATTACC R: GTGGCTCGCAAAGGACA	(AG) ₁₁	200-270	208-246	59	KX074068
SS2-15	F: FAM-TAGAGGAACAAAGAGTGGGAGA R: CGAGACTACGGGACAGACG	(CA) ₁₆	134-230	–	61	KX074069
SS3-1	F: FAM-TC TTTGAAGCCTTTTAGTC R: ACCGTCACAGAAGCCAC	(GCT) ₉	219-246	214-238	61	KX074070
SS4-1	F: FAM-GGAGAAGCCAGCCAAGC R: CACAGCCAGTCTGCCAC	(CAGG) ₁₁	102-186	103-131	61	KX074071
SS4-2	F: FAM-TGGGAATGAGGACACG R: TGAGGACCCCTGACAAGC	(CAGA) ₇	186-246	210-278	57	KX074072
SS4-4	F: FAM-TGTGACCCGAGGACTAAA R: GCGAGGGAAGGAAATGTG	(TAGA) ₃₇	215-319	–	53	KX074073
SS4-5	F: HEX-TGGTTATCAGTGTCCAGTTTT R: CAGAGGAGCATTAGCAGAG	(TTTC) ₈	241-309	–	56	KX074074
SS4-6	F: FAM-CAGTCAGAAGTGGCGTAT R: CGTCAGAAATGAATGGA	(CTTT) ₁₃	179-215	164-192	53	KX074075
SS6-1	F: FAM-TGTGAGTTGTCTGCTGTT R: ACTGGTTTTGATGTGGGTT	(TGAATC) ₅	202-232	–	58	KX074076

Ta = annealing temperature; “–” = not applicable.

Cross-species amplification of developed SSRs

Cross-species amplifications of polymorphic SSRs developed from *S. sihama* were tested in the closely related species *S. japonica* (Tables 1 and 2) using the amplification and genotyping method described above. Each locus with at least one band of the expected size was considered transferable.

Data analysis

The number of alleles (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E), and polymorphism information content (PIC) at each locus were calculated using Cervus

3.0.7 (Kalinowski et al., 2007). Tests for deviation from Hardy-Weinberg equilibrium (HWE) at each locus were carried out in GenePop 4.3 (Rousset, 2008) with the default settings. Significant P values were adjusted for multiple comparisons using the sequential Bonferroni correction ($\alpha = 0.05$) (Rice, 1989). Frequency of null alleles (*Fua*) was evaluated by Micro-Checker v.2.2.3 (Van Oosterhout et al., 2004) based on the Brookfield-1 method. In addition, the inbreeding coefficient (*Fis*) of Weir and Cockerham's version was calculated using the FSTAT 2.9.3 program package (Goudet, 2001).

Table 2. Characterization of 18 polymorphic microsatellite loci in *Sillago sihama* and 13 transferable loci in *Sillago japonica*.

Locus	<i>S. sihama</i>									<i>S. japonica</i>						
	Samples	N_A	H_O	H_E	PIC	HWE	<i>Fis</i>	<i>Fua</i>	Samples	N_A	H_O	H_E	PIC	HWE	<i>Fis</i>	<i>Fua</i>
SS2-1	31	17	0.839	0.895	0.870	0.0488	0.064	0.022	–	–	–	–	–	–	–	–
SS2-2	32	11	0.875	0.834	0.800	0.4079	-0.050	-0.030	24	9	0.625	0.807	0.759	0.0867	0.229	0.092
SS2-3	32	17	0.625	0.873	0.846	0.0000*	0.287	0.126	21	11	0.571	0.852	0.815	0.0007*	0.335	0.142
SS2-4	32	20	0.938	0.934	0.914	0.4934	-0.004	-0.009	23	15	0.435	0.916	0.887	0.0000*	0.531	0.243
SS2-6	32	14	0.938	0.873	0.845	0.1887	-0.075	-0.042	24	4	0.167	0.568	0.459	0.0002*	0.711	0.250
SS2-10	32	14	0.719	0.824	0.794	0.2946	0.130	0.051	24	16	0.625	0.915	0.887	0.0044	0.322	0.143
SS2-11	32	11	0.750	0.792	0.759	0.2793	0.053	0.017	24	17	0.875	0.866	0.838	0.8753	-0.010	-0.015
SS2-12	32	9	0.719	0.754	0.715	0.5648	0.048	0.014	24	17	0.875	0.922	0.895	0.9263	0.052	0.015
SS2-13	32	16	0.719	0.801	0.775	0.0528	0.104	0.039	24	18	0.958	0.929	0.903	0.3506	-0.032	-0.026
SS2-14	31	19	0.968	0.950	0.931	0.0681	-0.019	-0.017	24	19	0.583	0.950	0.926	0.0000*	0.391	0.180
SS2-15	32	18	0.531	0.907	0.884	0.0000*	0.418	0.191	–	–	–	–	–	–	–	–
SS3-1	32	10	0.719	0.837	0.804	0.1031	0.144	0.058	24	9	0.833	0.869	0.833	0.7973	0.042	0.009
SS4-1	32	17	0.875	0.917	0.895	0.1472	0.047	0.015	23	7	0.609	0.839	0.796	0.0359	0.279	0.116
SS4-2	32	8	0.844	0.803	0.761	0.1749	-0.052	-0.030	24	14	0.833	0.910	0.881	0.1128	0.085	0.030
SS4-4	32	24	0.875	0.944	0.925	0.0831	0.074	0.028	–	–	–	–	–	–	–	–
SS4-5	32	15	0.875	0.868	0.841	0.1207	-0.009	-0.011	–	–	–	–	–	–	–	–
SS4-6	31	5	0.226	0.319	0.301	0.0025*	0.295	0.067	23	3	0.130	0.335	0.295	0.0001*	0.616	0.149
SS6-1	32	6	0.500	0.573	0.528	0.1787	0.130	0.041	–	–	–	–	–	–	–	–
Mean		13.94	0.752	0.817	0.788		0.088			12.23	0.625	0.821	0.783		0.273	

N_A = number of alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; PIC = polymorphism information content; HWE = Hardy-Weinberg Equilibrium; *Fis* = inbreeding coefficient; *Fua* = frequency of null alleles; “–” = not applicable. Asterisks indicate significant departure from HWE after Bonferroni correction ($\alpha = 0.05$) for *S. sihama* ($P \leq 0.0028$) and *S. japonica* ($P \leq 0.0038$).

RESULTS

Quality assessment of Illumina sequencing

A total of 1632 Mb of paired-end raw sequencing data was generated using the Illumina Hiseq 2000 platform. After data filtering and error sequencing removal, 1363.48 Mb high-quality clean sequences were obtained for further analysis. The final assembled *S. sihama* genome was 853.48 Mb with a scaffold N50 of 1870 bp, and average GC content was 43.3%. The scatterplot of the GC content and sequencing depth showed a concentrated distribution, indicating that the assembled genome was free of contamination. The coverage of the contigs reached 90.92% with 9.51X sequencing depth.

Identification of genomic SSRs

Based on microsatellite searching, a total of 27,288 SSRs of longer than 22 bp were

identified in the *S. sihama* genome. Among them, more frequent repeats were dinucleotide SSRs (73.55%), followed by trinucleotide (9.87%), tetranucleotide (9.53%), hexanucleotide (3.89%), and pentanucleotide (2.94%). Although the longer repeat motifs (4-6 mers) shared relatively small proportions, the number of this type of motifs (4521) was considerable in the genome studied.

Polymorphism of detected SSRs

Of 133 primer sets tested, 88 pairs (66.2%) yielded specific PCR products and were selected to assess the polymorphism and PCR characterization of 32 individuals of *S. sihama*. In total, 18 SSR markers (20.5%) produced reliable PCR amplicons and were shown to be polymorphic, with allele sizes ranging from 102 to 348 bp. These 18 SSR sequences were submitted to the NCBI (GenBank accession Nos.: KX074059-KX074076, Table 1).

The N_A per locus varied from 5 to 24 (mean = 13.94, same as below, Table 2). H_O and H_E ranged from 0.226 to 0.968 (0.752) and from 0.319 to 0.950 (0.817), respectively. The PIC ranged from 0.301 to 0.931 with an average of 0.788. After Bonferroni correction (adjusted P value = 0.0028), significant deviations from HWE were detected at three loci (SS2-3, SS2-15, SS4-6), which revealed notable heterozygote deficiency. Meanwhile, the presence of both null alleles and inbreeding was detected at these three loci, with *Fua* and *Fis* ranging from 0.067 to 0.191 and from 0.287 to 0.418, respectively.

Cross-species amplification

Cross-species amplification results show that 13 of 18 loci were successfully amplified across 24 *S. japonica* individuals (Tables 1 and 2). The allele sizes of 13 loci varied between 103 and 297 bp. The values of N_A ranged from 3 to 19 (12.23), and the H_O and H_E ranged from 0.130 to 0.958 (0.625) and from 0.335 to 0.950 (0.821), respectively. The PIC values ranged from 0.295 to 0.926 with an average of 0.783. Eight loci conformed to the HWE after sequential Bonferroni correction (adjusted P value = 0.0038), while other 5 loci (SS2-3, SS2-4, SS2-6, SS2-14, and SS4-6) exhibited significant heterozygote deficiency. The deviated loci were simultaneously found to have null alleles and inbreeding, as indicated by the high positive values of both *Fua* (0.142-0.250) and *Fis* (0.335-0.711).

DISCUSSION

Characterization and mining of genomic SSRs in *S. sihama*

Based on NGS, a total of 853.48 Mb of clean genome was assembled with high coverage and sequencing depth, which provided high single-base accuracy and effective assembly for the *S. sihama* genome. This large genomic dataset produced 27,288 potential SSRs with various microsatellite motifs. Consistent with other fish genomes (Villanova et al., 2015), the majority of SSRs in *S. sihama* were shorter repeat motifs (22,763 SSRs of 2-3 mers), suggesting that it could be convenient and effective for isolating shorter motifs. Moreover, a considerable number of the longer repeat motifs (4521 SSRs of 4-6 mers) were also identified in the present study, whereas this type of SSRs are difficult to isolate using the traditional method. It has been demonstrated that longer motifs show more accuracy and efficiency of

allele assignment than shorter motifs (Restrepo et al., 2015). Thus, in the genomic era, it is necessary for researchers to develop longer SSR marker motifs rather than focusing on shorter motifs. Given these considerations, primer screening and polymorphism testing for longer motifs as well as shorter motifs were conducted in *S. sihama*.

Development of microsatellite markers and population genetic parameters for *S. sihama*

As a result, 18 polymorphic SSRs with four type motifs, including 11 di-, 1 tri-, 5 tetra- and 1 hexanucleotide repeats, were successfully isolated for *S. sihama*. The N_A per locus varied from 5 to 24 (13.94), with H_o ranging from 0.226 to 0.968 (0.752). These values were comparable to those reported in 32 other fishes (mean $N_A = 13.5$, mean $H_o = 0.66$) (DeWoody and Avise, 2000) but notably higher than those previously reported for *S. sihama* (N_A , 2-16, 6; H_o , 0.083-1.000, 0.631) (Guo et al., 2012), indicating considerable variability in these 18 SSR markers. According to the criteria defined by Botstein et al. (1980), 17 of 18 loci were highly informative ($PIC = 0.715-0.925 > 0.5$), and 1 locus (SS4-6) was reasonably informative ($0.25 < PIC = 0.301 < 0.5$). The presence of such highly polymorphic SSR markers (especially the six longest motifs) in the present study indicates that these loci can provide new insights for further research on the population genetic resources and conservation of *S. sihama*.

It has been reported that HWE deviation relative to heterozygote deficiency in microsatellites is commonly found in marine fishes (Yu et al., 2002; Lin et al., 2012). This deviation is usually caused by several major factors, including null alleles, inbreeding, the Wahlund effect, selection against heterozygotes, and population degradation (Yu et al., 2002). Our results show that the presence of both null alleles and inbreeding in three loci of *S. sihama* might cause heterozygote disequilibrium. With only one population tested, the Wahlund effect (subpopulations existing in samples) could not be considered as a reason for the HWE deviation in this study. However, other two possible causes (natural selection or population degradation) for HWE deviation could not be excluded in *S. sihama*, even though the tested population exhibited a high level of genetic diversity, as the natural resource of *S. sihama* has been significantly decreasing over the last 20 years (Lu et al., 2008; Liu et al., 2010). For further investigation, it is necessary to increase sample sizes and expand sample ranges to better understand the population evolution of *S. sihama*. In spite of this, the genetic parameter analyses of this study can provide particular information for the assessment of population genetic resources.

Transferability of the developed microsatellite markers to *S. japonica*

The success rate of cross-species amplification from *S. sihama* to *S. japonica* (13 of 18 loci, 72.2%) was much higher than that between other *Sillago* species; for instance, that of *S. japonica* to *S. parvisquamis* was 45.5% (Ueno et al., 2013) and that of *S. parvisquamis* to *S. japonica* was 63.6% (Umino et al., 2013). A significant negative correlation was found between cross-species amplification success and genetic divergence in the source species of SSRs (Addamo et al., 2015). Previous molecular studies demonstrate that the genetic distance between *S. sihama* and *S. japonica* is less than the distance between *S. japonica* and *S. parvisquamis* (Gao et al., 2011; Niu et al., 2016). Therefore, the high transferability of our SSR markers might result from the closer phylogenetic relationship between *S. sihama* and *S.*

japonica. In addition, the different types of SSRs developed in the current study may increase their transferability.

In the present study, the microsatellite diversity values ($N_A = 3-19$, 12.23; $H_O = 0.130-0.958$, 0.625) of *S. japonica* were slightly lower than those of *S. sihama* but higher than those reported for 13 freshwater fishes (mean $N_A = 9.1$, mean $H_O = 0.54$) (DeWoody and Avise, 2000). Moreover, the PIC values showed high levels of polymorphism information for most of the 13 transferable loci. These results suggest that our transferable loci exhibited a high level of genetic variation. As analyzed, due to the null alleles and inbreeding, the five transferable loci detected in *S. japonica* showed a significant departure from HWE, with two loci having a null allele frequency threshold of >0.2 (0.243 for SS2-4 and 0.250 for SS2-6). According to Dakin and Avise (2004), if the frequency of a null allele is >0.2 , the mean exclusion probability of parentage analysis will be significantly overestimated. For this reason, SS2-4 and SS2-6 should be ruled out for future parentage analysis or individual identification in *S. japonica*. It follows that the other 11 transferable loci have great potential to be used for genetic diversity and population structure studies in *S. japonica*. Because of the lack of other Sillaginidae species in our study, the transferability of the 18 developed loci to them is unknown. Thus, future study is needed to perform further cross-amplification testing in Sillaginidae fishes.

In this study, a stepwise approach including Illumina sequencing, primer screening, and SSR marker validation was first established to isolate SSR markers for sillaginid fishes. Using this method, we easily identified a large number of various SSRs from the genome and rapidly developed a set of polymorphic SSR markers for *S. sihama*. Moreover, the majority of isolated loci also showed high transferability to the closely related species *S. japonica*. The novel and high polymorphic SSRs (especially the longer motifs) discovered in this study will facilitate future estimation of genetic diversity, population structure and evolution, and molecular breeding for *S. sihama*. They also provide potential markers for *S. japonica*. Furthermore, the results indicate that our established approach is convenient and cost-effective, and is also suitable for SSR marker isolation of other sillaginid fishes.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research financially supported by the Special Fund for Agro-Scientific Research in the Public Interest of China (#201403008), the Fund of Key Laboratory of South China Sea Fishery Resources Exploitation & Utilization, Ministry of Agriculture, China (#FREU2015-05), the Project for Outstanding Young Teachers in Higher Education of Guangdong, China (#Yq2013093), and the National Natural Science Foundation of China (#31372532).

REFERENCES

- Abdelkrim J, Robertson B, Stanton JA and Gemmel N (2009). Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* 46: 185-192. <http://dx.doi.org/10.2144/000113084>
- Addamo AM, García-Jiménez R, Taviani M and Machordom A (2015). Development of microsatellite markers in the deep-sea cup coral *Desmophyllum dianthus* by 454 sequencing and cross-species amplifications in scleractinia order. *J. Hered.* 106: 322-330. <http://dx.doi.org/10.1093/jhered/esv010>

- Allentoft M, Schuster SC, Holdaway R, Hale M, et al. (2009). Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* 46: 195-200. <http://dx.doi.org/10.2144/000113086>
- Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27: 573-580. <http://dx.doi.org/10.1093/nar/27.2.573>
- Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- Boutin-Ganache I, Raposo M, Raymond M and Deschepper CF (2001). M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques* 31: 24-26, 28.
- Dakin EE and Avise JC (2004). Microsatellite null alleles in parentage analysis. *Heredity (Edinb)* 93: 504-509. <http://dx.doi.org/10.1038/sj.hdy.6800545>
- DeWoody JA and Avise JC (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J. Fish Biol.* 56: 461-473. <http://dx.doi.org/10.1111/j.1095-8649.2000.tb00748.x>
- Duan CX, Li DD, Sun SL, Wang XM, et al. (2014). Rapid development of microsatellite markers for *Callosobruchus chinensis* using Illumina paired-end sequencing. *PLoS One* 9: e95458. <http://dx.doi.org/10.1371/journal.pone.0095458>
- Gao TX, Ji DP, Xiao YS, Xue TQ, et al. (2011). Description and DNA barcoding of a new *Sillago* species, *Sillago sinica* (Perciformes: Sillaginidae), from coastal waters of china. *Zool. Stud.* 50: 254-263.
- Goudet J (2001). FSTAT: a program to estimate and test gene diversities and fixation indices (Version 2.9.3). Available at [<http://www.unil.ch/izea/softwares/fstat.html>].
- Grover A and Sharma PC (2016). Development and use of molecular markers: past and present. *Crit. Rev. Biotechnol.* 36: 290-302. <http://dx.doi.org/10.3109/07388551.2014.959891>
- Guo YS, Wang ZD, Yan CZ, Zhang YL, et al. (2012). Isolation and characterization of microsatellite DNA loci from *Sillago sihama*. *J. Genet.* 91: e32-e36.
- Huang Y, Du T and Huang HL (2013). A study on artificial breeding of *Sillago sihama* Forskål. *J. Guangdong Ocean Univ.* 33: 15-21.
- 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. <http://dx.doi.org/10.1111/j.1365-294X.2007.03089.x>
- Li R, Zhu H, Ruan J, Qian W, et al. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20: 265-272. <http://dx.doi.org/10.1101/gr.097261.109>
- Lin L, Zhu L, Liu SF, Tang QS, et al. (2012). Polymorphic microsatellite loci for Japanese Spanish mackerel (*Scomberomorus niphonius*). *Genet. Mol. Res.* 11: 1205-1208. <http://dx.doi.org/10.4238/2012.May.8.2>
- Liu D, Guo Y, Wang Z and Liu C (2012). Phylogenetics inferred from mitogenome and control region of Silver Sillago, *Sillago sihama*. *Mitochondrial DNA* 23: 255-263. <http://dx.doi.org/10.3109/19401736.2012.674118>
- Liu JD, Zhu LX, Lu HS, Zhu JX, et al. (2010). Estimation of growth and mortality parameters of the silver sillago (*Sillago sihama*) in Beibu Gulf. *J. Zhejiang Ocean Univ. (Nat. Sci.)*. 29: 64-69. (In Chinese with English abstract)
- Lu ZB, Chen X and Du JG (2008). The population dynamics and parameter of growth and mortality of *Sillago sihama* in the Minnan-Taiwan fishing grounds. *Mar. Fish. Res.* 29: 47-53.
- Ma H, Cui H, Ma C and Ma L (2012). High genetic diversity and low differentiation in mud crab (*Scylla paramamosain*) along the southeastern coast of China revealed by microsatellite markers. *J. Exp. Biol.* 215: 3120-3125. <http://dx.doi.org/10.1242/jeb.071654>
- Niu S, Wu R, Liu Y and Wang X (2016). The High-throughput sequencing of *Sillago japonica* mitochondrial genome reveals the phylogenetic position within the genus *Sillago*. *Mitochondrial DNA A DNA Mapp Seq Anal* 27: 3815-3816.
- Page RB, Sankamethawee W, Pierce AJ, Sterling KA, et al. (2014). High throughput sequencing enables discovery of microsatellites from the puff-throated bulbul (*Alophoixus pallidus*) and assessment of genetic diversity in Khao Yai National Park, Thailand. *Biochem. Syst. Ecol.* 55: 176-183. <http://dx.doi.org/10.1016/j.bse.2014.03.032>
- Restrepo A, Páez VP, Vásquez A and Daza JM (2015). Rapid microsatellite marker development in the endangered neotropical freshwater turtle *Podocnemis lewyana*, (Testudines: Podocnemididae) using 454 sequencing. *Biochem. Syst. Ecol.* 59: 220-225. <http://dx.doi.org/10.1016/j.bse.2015.01.017>
- Rice WR (1989). Analyzing tables of statistical tests. *Evolution* 43: 223-225. <http://dx.doi.org/10.2307/2409177>
- Rousset F (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103-106. <http://dx.doi.org/10.1111/j.1471-8286.2007.01931.x>
- Sambrook J and Russell DW (1989). Molecular cloning: A laboratory manual. 3rd edn. Cold Spring Harbor Laboratory Press, New York.
- Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18: 233-234. <http://dx.doi.org/10.1038/72708>

- Schuster SC (2008). Next-generation sequencing transforms today's biology. *Nat. Methods* 5: 16-18. <http://dx.doi.org/10.1038/nmeth1156>
- Shao KT (2016). Taiwan Fish Database. Available at [<http://fishdb.sinica.edu.tw>].
- Ueno K, Watanabe M, Ahmad-Syazni K, Koike M, et al. (2013). Eleven novel microsatellite loci for Japanese whiting (*Sillago japonica*) and cross amplification in the endangered small-scale sillago (*Sillago parvisquamis*). *Conserv. Genet. Resour.* 5: 659-662. <http://dx.doi.org/10.1007/s12686-013-9876-x>
- Umino T, Ueno K, Mihara T, Koike M, et al. (2013). Isolation of eleven polymorphic microsatellite loci for the endangered sillago parvisquamis and cross-species amplification with *Sillago japonica*. *Conserv. Genet. Resour.* 5: 771-773. <http://dx.doi.org/10.1007/s12686-013-9904-x>
- Van Oosterhout C, Hutchinson WF, Wills DP and Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4: 535-538. <http://dx.doi.org/10.1111/j.1471-8286.2004.00684.x>
- Villanova GV, Vera M, Díaz J, Martínez P, et al. (2015). Isolation and characterization of 20 polymorphic microsatellite loci in the migratory freshwater fish *Leporinus obtusidens* (Characiformes: Anostomidae) using 454 shotgun pyrosequencing. *J. Fish Biol.* 86: 1209-1217. <http://dx.doi.org/10.1111/jfb.12632>
- Wang L, Song N and Gao T (2014). Isolation and characterization of 24 polymorphic microsatellite loci in Japanese sillago (*Sillago japonica*). *Conserv. Genet. Resour.* 6: 581-584. <http://dx.doi.org/10.1007/s12686-014-0145-4>
- Yang T, Fang L, Zhang X, Hu J, et al. (2015). High-throughput development of SSR markers from pea (*Pisum sativum* L.) based on next generation sequencing of a purified Chinese commercial variety. *PLoS One* 10: e0139775. <http://dx.doi.org/10.1371/journal.pone.0139775>
- Yu HT, Lee YJ, Huang SW and Chiu TS (2002). Genetic analysis of the populations of Japanese anchovy (Engraulidae: *Engraulis japonicus*) using microsatellite DNA. *Mar. Biotechnol.* 4: 471-479. <http://dx.doi.org/10.1007/s10126-002-0035-8>
- Zhang HR, Niu SF, Wu RX, Zhai Y, et al. (2016). Development and characterization of 26 polymorphic microsatellite markers in *Lateolabrax maculatus* and cross-species amplification for the phylogenetically related taxa. *Biochem. Syst. Ecol.* 66: 326-330. <http://dx.doi.org/10.1016/j.bse.2016.05.008>