ONLINE RESOURCES

Microsatellite development and characterization for *Saurogobio dabryi* Bleeker, 1871 in a Yangtze river-connected lake, China

HONG GAO LIU*, ZHI YANG, HUI YUAN TANG, YUN GONG and LI WAN

Key Laboratory of Ecological Impacts of Hydraulic Projects and Restoration of Aquatic Ecosystem of Ministry of Water Resources, Institute of Hydroecology, Ministry of Water Resources and Chinese Academy of Sciences, Wuhan, Hubei 430079, People's Republic of China

[Liu H. G., Yang Z. Tang H. Y., Gong Y. and Wan L. 2017 Microsatellite development and characterization for *Saurogobio dabryi* Bleeker, 1871 in a Yangtze river-connected lake, China. *J. Genet.* **96**, e1–e4. Online only: http://www.ias.ac.in/jgenet/OnlineResources/96/e1.pdf]

Introduction

Saurogobio dabryi Bleeker, 1871 (Chinese lizard gudgeon) is a diploid fish ($2n = 50$), belongs to the family Cyprinidae, subfamily Gobioninae (Tang *et al.* 2011) and distributed in the east Asia (Dai and Yang 2002). As an indispensable component of the food web, it is essential for the stability and integrity of the Yangtze ecosystem (Zhang *et al.* 2015). Moreover, it is an excellent species for landscape genetic and ecological researches because of their characteristics of migration, small size, fast reproduction and susceptibility to anthropogenic disturbances (Ru and Liu 2013). However, so far it has no genome information and very few markers have been published (Yu *et al.* 2009). To develop massive *S. dabryi* microsatellite markers, we performed an enriched microsatellite library sequencing (EMLS) and a genome sequencing (GS) methods based on 454 GS-FLX Titanium system.

Materials and methods

Twenty-seven fish specimens were collected from the Yangtze river-connected Shijiu lake (118°45′4.19″N and 31°29′26.98″E) by gillnetting in November 2014. Genomic DNA was prepared from fin by phenol–chloroform extraction method (Sambrook *et al.* 1989). In EMLS method, microsatellite-enriched libraries were constructed using two Specimens following the fast isolation by AFLP of sequences containing repeats protocol (FIASCO) (Rikalainen 2013) with modifications. One microgram genomic DNA was digested by 10 U *Msp*I, ligated to 100 pmol labelled sequencing adapters by 200 U T4 DNA ligase, smoothed by 4 U T4 DNA polymerase, and then amplified by polymerase chain reaction (PCR) using sequencing primers. Amplicons were hybridized with 42 designed 5′-biotinylated probes that cover all dinucleotide, trinucleotide and tetranucleotide motifs. The probes consisted of $(AB)_{15}$, $(AVB)_9$ and $(AVNB)_7$ (B is C, G or T; V is A, C or G). Microsatellite-containing amplicons were trapped by 150 μL activated streptavidin-coated magnetic beads (Promega, Madison, USA) and then eluted. The eluent was carefully collected, mixed and sequenced on a quarter-plate for the 454 GS FLX+ sequencer following the manufacturer's manual (Roche Diagnostics, Mannheim, Germany). Adapters and short sequences (<200 bp) were removed from raw reads using Newbler software (ver. 2.9, Roche Diagnostics). In GS method, the 454 FLX+ genome libraries were constructed using genomic DNA of two specimens and labelled sequencing adapters, then mixed together and sequenced on a quarter-plate for the 454 GS FLX+ sequencer. The reads were coarsely assembled using the Newbler software. In both the methods, microsatellite loci were searched *in silico* according to the criteria: dinucleotide motifs with ≥6 repeats, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide motifs with ≥4 repeats. Primers were designed using Primer 3 software (Untergasser *et al.* 2012).

Twenty-seven accessions were genotyped at 40 microsatellite loci. Forward primers were labelled with fluorescent dye (FAM). PCR mixture (15 μL) consisted of ~30 ng genomic DNA, 7.5 μL 2 × Power Taq PCR MasterMix (Tiangene, Beijing, China) and 0.3 μM of each primer. PCRs were performed on a thermal cycler MyCycler (Bio-Rad, Hercules, USA). PCR conditions were: 95°C 5 min; 94°C 30 s, 58°C 30 s, 72°C 30 s, 28 cycles; 72°C 5 min. Two microlitres of diluted product were added to 7.75 μL Hi Di^{TM} formamide and 0.25 μL Gene Scan-500 LIZ^{TM} size standard, heated at 94°C for 5 min and immediately chilled

*For correspondence. E-mail: hugolaugene@gmail.com.

Keywords. microsatellites; 454 pyrosequencing; Yangtze river; *Saurogobio dabryi.*

Table 1. Characterization of microsatellite markers for Sharpbelly population in Shijiu lake.

Locus name	Primer sequences (5′–3′)	Repeat motif	Locus size (bp)	Sample size	N_A	H_O	H_E	PIC	GenBank accession
*sdF5010***	F: GCTCCATTCCCAACCTATAACACA R: TGCCTTGATGAACGGTTATGTATG	ACAT	78–109	24	10	0.63	0.86	0.89	KX250317
sdF5117	F: CGAGTAGACTTATGCGTTTCAATGG R: GAATTTCCCCCACTGCTAATACAA	AAT	134–153	27	8	0.7	0.81	0.56	KX250318
*sdF5129***	F: TTACAGGCCAGTGATGTATGTTGC R: CTTGAATGAATGAACGATCCTGTG	TC	123–158	27	19	0.74	0.93	0.82	KX250319
sdF5145	F: AACTGTAGGGCACGACAAATTGAT R: AGTCTAAACCCGTCTGCAAGAATG	AGAC	188–260	27	18	0.85	0.93	0.5	KX250320
sdF5163	F: TCTCAGATGACGTTGAGCATATTGA R: CATTCATCTGGGCTCACTAAAACA	ATCT	158–210	26	15	0.81	0.91	0.89	KX250321
sdF5180	F: TGTGGGTCTGAGATTTGTGATGTT R: GGCCCATGCGTAAGTAGGTTATC	AG	127–143	27	9	0.74	0.81	0.66	KX250322
*sdF5226***	F: AGACTCGGCTGCACCATCTATTAC R: AGCAATTTGGGAGATTTGTGATGT	AGTT	113–151	20	10	0.55	0.86	0.49	KX250323
*sdF5234***	F: TTTGAGACTGGCTCAGATGTTCAG R: CAGAGTGGAAATCTGGGTAAATGG	GCT	158–168	27	8	0.67	0.83	0.89	KX250324
*sdF5257***	F: TGAGGAAGAGGAGGAGACTTCAGA R: CTCTATGAGCTCAGATGGGGATTC	GAG	125–134	27	4	0.26	0.41	0.27	KX250325
*sdF5261***	F: ATGACAAAGGAAAACCACGAGAAA R: TGCAGATACAGCGCATCACTTTAG	AAAC	114–145	26	16	0.81	0.87	0.71	KX250326
sdF5271	F: CGAGTAGACGGAGTTCAGCACAT R: TGTTGTATAAGCGTGTGATGTGGA	AC	122–124	27	2	0.74	0.48	0.88	KX250327
*sdF5297***	F: GCATGTTGACCTGCAAGATAGAAA R: AACATAAATGAAGCCTTGGTGAGC	TA	118–131	27	3	0.15	0.14	0.92	KX250328
sdF5302	F: CGCTGGATATTGAAAGAGTCTGGT R: CTTGTCCTCCTAATCCTCCTCCTC	GGA	160–172	27	9	0.63	0.77	0.91	KX250329
*sdF5318***	F: GTACGTGGTGTATTGCCAGAACAA R: CCAAAACAGTCCCTCATGCTTAAT	AC	231–261	26	11	0.46	0.54	0.53	KX250330
sdF5328	F: GGTGGTGATTAGTAGGCTGGTGAC R: CTGACTGCCCCCAATTACAGAATA	TAT	108–116	27	4	0.26	0.65	0.29	KX250331
*sdF5340***	F: CGAGTAGACTAGACGCCCACACTT R: TTCACAATCCACATTTCCTTAAACC	CTTT	126–210	26	20	0.85	0.81	0.81	KX250332
sdF5351	F: TTTCTGTACTTGTTAGTTTGGGGTCA R: GATCAATTAAATGCATCATTGCTGA	TTTC	100–109	27	6	0.37	0.65	0.77	KX250333
*sdF5395***	F: AAGAAGAGGAAGGAGAAGGCGAG R: CGGTTTCCTCCTGTTTTTCTCAGT	GAG	74–99	27	10	0.81	0.85	0.39	KX250334
sdF5432	F: TTCGGTTTTTACCGTGATCTCAAT R: GATAACACTGACTGACTGCTGGGA	GAG	139–147	26	5	0.35	0.74	0.23	KX250335
*sdF5433***	F: AAAAGACGCCCACACTTTTATCTG R: TCACAATCCACATTTTCCTTAAACC	CTTT	115–333	27	6	0.15	0.7	0.84	KX250336
*sdF5439***	F: CAGAGCCACATATAAGCACACCAC R: CGGTTGTACGAAGACAGACTCAGA	TCG	135–153	26	7	0.35	0.41	0.22	KX250337
sdF5466	F: CACCAGAGCCACATATAAGCACAC R: ACGAAGACAGGACTCCAGAGAAGA	TCG	121–129	27	6	0.3	0.54	0.65	KX250338
sdF5484	F: TTTCCGCAGGCAGAGATAGTAAAG R: GAGCAGAGTGGCGTTAGAGGG	AC	135–170	25	17	0.72	0.91	0.39	KX250339
sdF5506	F: TCAAATGTATTTCAGCCGATGCTA R: ATGAGGGTGTTTGTGTATGTGGGT	ACT	131–147	27	6	0.52	0.62	0.62	KX250340
sdF5513	F: TAAGCACACCAGTCCCATTCACTA R: ACAGACTCAGGACTGGAGACGAAG	ATC	94–167	26	9	0.62	0.69	0.8	KX250341
sdF6119	F: TCATGCAAACAAGAGTTGAGCAAT R: GGCTCTCAAAATTCTCCCAAGAAG	TGATT	76–93	27	6	0.7	0.73	0.04	KX250342
sdF6273	F: CACAGACGAGGATTCCAGTCG R: ACATGCAGTCCCTTACCTGTTCTC	GAG	156–174	27	5	0.81	0.76	0.83	KX250343
*sdF6293***	F: CGTCTAGTTTTGTCGCTCTGTTGA R: TGCTTACTTGAGGTATGCTGGTTG	CATA	78–128	27	8	0.3	0.77	0.26	KX250344
*sdF6322**	F: CTTAACTAAACCGCACCCCTCTCT R: GTCTCCTCTCCTCTGAGCTGACTG	TCC	148–149	27	2	0	0.07	0.4	KX250345
*sdF6324**	F: TGTAGCGTTCCAAACCTTTCTCAT R: GGGAGTATGGTGGGGAAGTCTTAT	CCT	136–137	25	2	0	0.08	0.23	KX250346

Table 1 (*contd*)

Locus name	Primer sequences (5′–3′)	Repeat motif	Locus size (bp)	Sample size	N_A	H_O	H_E	PIC	GenBank accession
sdF6400	F: GTGGCTCCTTCCCTTTCACAG R: AGCAGATTGTACATGTGGCTCAAG	GGA	101–131	26	9	0.81	0.84	0.68	KX250347
sdF6432	F: TTCTGGACATTCCACCTAAAACATC R: TCATAGAGGAAACAATGTGTAATCCAA	CTATT	135–198	26	17	0.88	0.93	0.44	KX250348
sdF6476	F: CGTCTAGTGCTGAAGGAGGTGAGT R: TCTCAGCCTGGAACACAGAGAGAT	ACTC	144–268	26	25	0.85	0.95	0.56	KX250349
*sdF6479***	F: TCCGTTGTTTAGGCTACTGATCAAA R: TGAGATGACATGACGATAGCTGTG	GATA	166–227	25	17	0.8	0.94	0.53	KX250350
*sdF7038***	F: GGCAAGCTGTTTGACTTATTCATTC R: TTTAACCCACTGGACCACCATATC	AGT	122–450	27	5	0.07	0.64	0.87	KX250351
sdF7068	F: TTAAAGCAGGGTTCACATCCAC R: ACGTGTGACATTTTTATCTAGGAATACA	AC	98–102	24	4	0.17	0.16	0.84	KX250352
*sdF7209***	F: ACACTACTCGTCTGCCGCAA R: ACAATGCAGATTGTTTCAAAGCAG	AAAT	98–102	27	2	1	0.51	0.63	KX250353
sdF7235	F: ATCTTTCTCAAAATCTGGCAGGTG R: AGTTGCTGTAGTGACCCAAAGAGC	TCA	129–213	27	4	0.22	0.21	0.66	KX250354
sdF8049	F: AAAACACACAAGCCAGAGAGACAG R: AGGTTGTTTGGTGTAAACGACGAT	ATC	75–94	20	4	0.3	0.35	0.55	KX250355
*sdF8133***	F: GTATACCCCCATGGGAAGAAGAAG R: GGGGGACTACTTTGTAAAAATGGC	AAT	116–131	27	5	0.81	0.68	0.23	KX250356

*Significant deviation from HWE ($P < 0.05$); **extremely significant deviation from HWE ($P < 0.01$). N_A, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity.

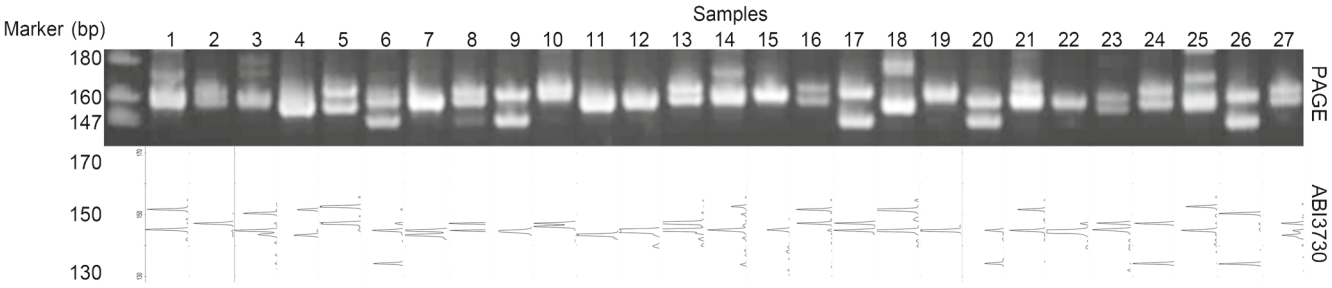

Figure 1. A representative genotyping profile. Twenty-seven *S. dabryi* accessions were genotyped at *sdF5117* locus on an ABI 3730 Genetic Analyzer and alleles were scored using GeneMapper 4.0. Results were consistent with those by 10% polyacrylamide gel electrophoresis (PAGE) and Gelred staining.

in ice for 2 min. Genotyping was carried out on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, USA) and alleles were scored using GeneMapper 4.0 (Applied Biosystems). Genetic diversity, including expected heterozygosity (H_E), observed heterozygosity (H_O), number of alleles (N_A) and polymorphism information content (PIC) were estimated using Arlequin 3.5.2.2 software (Excoffier and Lischer 2010) and Microsatellite toolkit (Park 2001). Possible deviations from Hardy–Weinberg equilibrium (HWE) were tested by Fisher's exact test with Bonferroni correction using the Arlequin software. The genotyping results were randomly examined by nonfluorescent PCRs, followed by 10% polyacrylamide gel electrophoresis (PAGE) and Gelred staining (Biotium, Hayward, USA).

Results and discussion

Using EMLS method, 8634 microsatellite loci were isolated from 24,281 reads (9.34 M nucleotides). Primers were cursorily designed for 6191 loci, of which 939 were predicted to be highly reliable (table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/). There were 170 motifs obtained, of which the most frequent was GT/CA (428, 18.43%), followed by AC/TG (406, 17.48%), AG/TC (138, 5.94%), AAT/TTA (123, 5.30%), GA/CT (101, 4.35%) and GTA/CAT (88, 3.79%). Dinucleotide motifs were the most frequent (4322, 45.81%), followed by trinucleotide (2416, 25.61%), tetranucleotide (2481, 26.30%), pentanucleotide

(80, 0.85%) and hexanucleotide motifs (135, 1.43%). To our surprise, pentanucleotide and hexanucleotide microsatellites appeared despite no related probe was designed.

Using GS method, 2332 microsatellite loci were isolated from 2115 contigs and 38,239 singletons (11.67 M nucleotides). Primers were cursorily designed for 1922 loci, of which 610 were predicted to be highly reliable (table 1 electronic in supplementary material). There were 117 motifs obtained, the most frequent was AG/TC (1552, 16.45%), followed by GT/CA (1067, 11.31%), AC/TG (1034, 10.96%), AAT/TTA (738, 7.82%), AGC/TCG (674, 7.14%) and GA/CT (641, 6.79%). Dinucleotide motifs were the most frequent (1132, 48.54%), followed by trinucleotide (601, 25.77%), tetranucleotide (281, 12.05%), pentanucleotide (48, 2.06%) and hexanucleotide motifs (270, 11.58%).

Genotyping results are shown in table 1. The average N_A per locus was 8.83 ranging from 2 to 25. The H_O and H_E were 0–1 and 0.07–0.95 with averages of 0.54 and 0.66, respectively. The H_E of 21 loci (52.50%) were $\geq$0.70 and the H_O of 15 loci (37.50%) were $\geq$0.7, indicating a high polymorphism and discriminability. The mean PIC was 0.62 ranging from 0.07 to 0.93. Twenty-nine loci (72.50%) had PICs at high polymorphism ($\geq$0.5) and six loci (15.00%) had PICs at moderate polymorphism (0.25–0.5). Except two loci, which had significant departure from HWE ($P < 0.05$), 17 loci showed extremely significance ($P < 0.01$) and 21 loci validated to be highly informative. At a randomly picked *sdF5117* locus (figure 1), our genotyping results using ABI 3730 method were confirmed by those using PAGE method.

Microsatellite markers are very popular in fish research for their high polymorphism, good reproducibility, wide distribution in the genome and codominant nature, and their development has been accelerated by next-generation sequencing platforms in the last decade (Sundaray *et al.* 2016). Taking advantage of both high throughput and long read length, 454 pyrosequencing has been one of the excellent choices (Kim *et al.* 2016). In our study, both GS and EMLS methods were verified as useful methods to develop fish microsatellites, and as expected, the EMLS method obtained a higher yield by microsatellite enrichment. Results would facilitate landscape genetic and ecological research, and shed light on fish conservation in the Yangtze basin.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (grant nos. 51179109, 51209151 and 51379134).

References

Dai Y. and Yang J. 2002 Description of a new subspecies of the genus *Saurogobio Bleeker* (1870). *Zool. Res.* **23**, 306–310.

Excoffier L. and Lischer H. E. 2010 Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* **10**, 564–567.

Kim K. S., Noh C. H., Moon S. J., Han S. H. and Bang I. C. 2016 Development of novel tetra- and trinucleotide microsatellite markers for giant grouper *Epinephelus lanceolatus* using 454 pyrosequencing. *Mol. Biol. Rep.* **43**, 541–548.

Park S. 2001 The Excel microsatellite toolkit. Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. thesis, University of Dublin, Dublin, Ireland.

Rikalainen K. 2013 Fast isolation by AFLP of sequences containing repeats. In *Methods in molecular biology* (ed. Stella K. Kantartzi), vol. 1006, pp. 57–66. Springer, New York, USA.

Ru H. J. and Liu X. Q. 2013 River–lake migration of fishes in the Dongting lake area of the Yangtze floodplain. *J. Appl. Ichthyol.* **29**, 594–601.

Sambrook J., Fritsch E. F. and Maniatis T. 1989 *Molecular cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sundaray J. K., Rasal K. D., Chakrapani V., Swain P., Kumar D., Ninawe A. S. *et al.* 2016 Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. *Aquacult. Int.* **24**, 1089–1102.

Tang K., Agnew M., Chen W., Vincent H. M., Raley M., Sado T. *et al.* 2011 Phylogeny of the gudgeons (Teleostei: Cyprinidae: Gobioninae). *Mol. Phylogenet. Evol.* **61**, 103–124.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., Remm M. *et al.* 2012 Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115.

Yu D., Ding B. and Liu H. 2009 Isolation and characterization of 11 polymorphic microsatellite loci for the *Saurogobio dabryi* (Teleostei: Cyprinidae). *Conserv. Genet. Resour.* **1**, 471–473.

Zhang X., Yu D., Wang H., Wan A., Chen M., Tao F. and Song Z. 2015 Effects of fish community on occurrences of Yangtze finless porpoise in confluence of the Yangtze and Wanhe Rivers. *Environ. Sci. Pollut. Res. Int.* **22**, 9524–9533.

Received 28 March 2016, in final revised form 17 May 2016; accepted 26 May 2016
Unedited version published online: 30 May 2016
Final version published online: 4 January 2017

Corresponding editor: RAJIVA RAMAN