Acta Herpetologica 14(2): 153-158, 2019
DOI: 10.13128/a_h-7755

AT-rich microsatellite loci development for *Fejervarya multistriata* by Illumina HiSeq sequencing

Yan-Mei Wang, Jing-Yi Chen, Guo-Hua Ding*, Zhi-Hua Lin

*ADI, College of Ecology, Lishui University, Lishui 323000, Zhejiang, China. * Correspondence author.* Email: guwoding@qq.com

Submitted on 2018, 22th May; Revised on: 2019, 3rd June; Accepted on: 2019, 23rd August
Editor: Emilio Sperone

Abstract. In our study, a total of 2561 sequences that contained microsatellite loci were found potentially to be used for primer design. Furthermore, Illumina HiSeq sequencing technology identified trinucleotide repeats and AT-rich repeats with the the highest proportion in our genomic DNA sequence library of *Fejervarya multistriata*. Eighteen new microsatellite loci of *F. multistriata* were isolated and we characterize these loci genotyping 48 individuals sampled from 3 populations in Lishui City, Zhejiang Province, China. Seventeen loci were polymorphic, with the number of alleles ranging from 2 to 11 within each population. The polymorphic information content, observed and expected heterozygosity ranged 0-0.845, 0-1.0 and 0-0.871, respectively. None of the loci was observed in linkage disequilibrium. One locus (FMA294) was deviated from Hardy-Winberg equilibrium in each population separately and combined. These informative microsatellite loci will be applicable for conservation genetic studies of *F. multistriata* across varying scales from inter-individual to inter-population.

Keywords. *Fejervarya multistriata*, genome, microsatellite, next-generation sequencing, polymorphism.

INTRODUCTION

Microsatellite DNA loci, also known as simple sequence repeats, occur at thousands of locations within the eukaryotic genome, and are highly variable and sufficient in nuclear genome (Ellegren, 2004; Wei et al., 2015; Shao et al., 2017). Therefore, microsatellite DNA loci are widely applied as molecular markers in population genetics, species identifying, genetic breeding and genetic map (Selkoe and Toonen, 2006; Abe et al., 2012; Wambulwa et al., 2016; Soulard et al., 2017). Thanks to the development of next-generation sequencing technology, both throughput and efficiency of developing microsatellite DNA has increased with a decreased cost of sequencing process. In recent years, microsatellite DNA markers have been quickly developed in many species at a low cost when using the next-generation sequencing technology on Illumina HiSeq and Ion Torrent PGM platforms (Yang et al., 2012; Lü et al., 2013; Zhang et al., 2013; Sultana et al., 2014; Igawa et al., 2015; Song et al., 2017).

Fejervarya multistriata (Anura: Dicroglossidae) is a species of frog, which is widely found in south of the Yellow River in China and some countries (regions) in Southeast Asia, such as northern India, Vietnam and Myanmar (AmphibiaChina, 2018). The conservation status of this species is listed as data deficient in IUCN (AmphibiaChina, 2018). This species prefers to inhabit paddy field and still water and its ovulation remains active from April to September every year (AmphibiaChina, 2018). As a dominant amphibian species, *F. multistriata* plays an important role in farmland ecosystem, and its population density in field has decreased due to urbanization (Li et al., 2016). Meanwhile, environmental degradation also threatens the survival of this species (Othman et al., 2009). In previous studies, mitochondrial D-loop sequences were used as molecular mark-

ISSN 1827-9635 (print)
ISSN 1827-9643 (online)

ers to study phylogeography of *F. multistriata* populations (Zhong et al., 2008). Twenty-one microsatellite loci, mainly including dinucleotide repeats, had been isolated for the species, and only approximately 24% loci had AT repeats (Chen et al., 2018). Here, we sequentially developed 18 new microsatellite DNA loci containing AT-rich repeats for *F. multistriata* by Illumina HiSeq sequencing. These new AT-rich microsatellite loci definitely would be useful in examining genetic diversity and protecting genetic resources of *F. multistriata*.

MATERIAL AND METHODS

Forty-eight *F. multistriata* individuals (n = 16 for each population) used in this study were collected by hand and net from 3 localities in Lishui City, Zhejiang Province, China, which were Lanshantou (LST, 119.7607°E, 28.36366°N), Baimashan (BMS, 119.1337°E, 28.63823°N) and Jiulongshan (JLS, 118.8452°E, 28.39538°N), respectively. Our experimental procedures are compliant with current laws on animal welfare and research in China, which are also specifically approved by the Animal Research Ethics Committee of Lishui University (Permit No. AREC-LU 2017-04).

Genomic DNA was extracted from toe muscle tissue of one male *F. multistriata* from LST population using the DNeasy Tissue Kit (Qiagen). The concentration of DNA sample was measured by using a spectrophotometer at 260 and 280 nm and DNA sample was quantified on an agarose gel. A 200-400 bp sequencing library was constructed according to the manufacturer instructions (Illumina). This library was sequenced using an Illumina HiSeq 2500 Platform with RAD-Tag at Novogene Bioinformatics Technology Co., Ltd (Beijing, China, http://www.novogene.com/). The microsatellite primer pairs of *F. multistriata* were designed using Primer Premier 3.0 software, which was used to check against potential primer dimers, hairpin structures and the occurrence of mismatches. Parameters for designing the primers were set as follows: primer length ranged from 18 bp to 24 bp with 22 as the optimum; PCR product size ranged from100-280 bp; melting temperature ranged from 50 °C to 65 °C with 55 °C as the optimum annealing temperature; GC content ranged from 30% to 70% with 50% as the optimum. Finally, thirty newly designed primer pairs were selected to synthesize and initially screened for microsatellite loci using total genomic DNA isolated from 6 *F. multistriata* individuals collected from the LST population.

PCR amplification reactions were performed using a thermal cycler (T100, Bio-Rad, USA). The total volume of each PCR mixture was 20 μL, containing 1 μL genomic DNA (100 ng/μL), 10 μL Premix Taq (TaKaRa, Japan), 1 μL of each primer (10 μM) and 6 μL double distilled H_2O. The conditions of the PCR amplification were as follows: 95 °C for 5 min, then 35 cycles at 95 °C for 30 s, T_a (the optimal annealing temperatures, see Table 1) for 30 s, 72 °C for 30 s, and a further extension at 72 °C for 10 min. Twenty-six primer pairs were further selected due to successful amplification in the 6 individuals, and the forward primer was labeled with FAM or HEX fluorescent dye (Sangon Biotech Ltd. Co., Shanghai, China). The PCR products were genotyped on an ABI 3730 sequencer (Applied Biosystems) and following data were analyzed with GeneMarker v1.8 software.

Population genetic parameters for polymorphic loci such as number of alleles (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E), polymorphic information content (PIC), P values in Hardy-Weinberg equilibrium (HWE) tests and linkage disequilibrium were calculated by Genepop 4.0 (Rousset, 2008), Cervus 2.0 (Marshall et al., 1998) and Fstat 2.9.3.2 (Goudet, 1995), respectively.

RESULTS AND DISCUSSION

Sequence data from Illunima HiSeq

We obtained a total of 6970707900 bp and 23235693 reads in a single sequencing run on Illunima HiSeq™ using RAD-Tag. The distribution frequency of read length for this species had a single peaks at approximately 125 bp. In a total of 307793 reads with more than 125 bp, 2561 reads contained microsatellite loci (0.83%).

Compared to a traditional library-based approach such as magnetic beads enrichment (Guo et al. 2015; Chang et al. 2016), next-generation sequencing technology is a more powerful approach to develop microsatellite markers due to its efficiency and low cost. Sufficient microsatellite sequences can be constructed in a genomic DNA sequence library on Illunima HiSeq™ using RAD-Tag. This result agrees with recent studies on other anuran species (Wei et al., 2015; Shao et al., 2017). Furthermore, microsatellite obtained rate maybe higher on Illunima HiSeq platform (e.g., 0.83% in our study) than on Ion Torrent PGM platform (e.g., 0.32–0.57% in Igawa et al., 2015).

Frequency and distribution of microsatellite loci in the genome

The length of the microsatellite loci ranged from 12 to 33 bp (15.7 ± 5.2, mean ± SD). The microsatellite DNA loci included 5 motif types: dinucleotide repeats (36.20%), trinucleotide repeats (52.60%), tetranucleotide repeats (9.64%), pentanucleotide repeats (0.90%) and hexanucleotide repeats (0.66%) (Fig. 1A). The frequency distribution of the 5 motif types was different significantly (G-test, G = 3085.9, df = 4, $P < 0.001$). The motif repeat number of microsatellite loci ranged from 4 to 16, while 97.66% of the microsatellite loci had 4-12 motif repeats, and motifs with more than 12 repeats were only with a frequency of <1.0% (Fig. 1B). There were 4 dinucleotide motif types, and the main types were AC/GT (50.70%), AT (35.17%) and AG/CT (13.92%) (Fig. 1C),

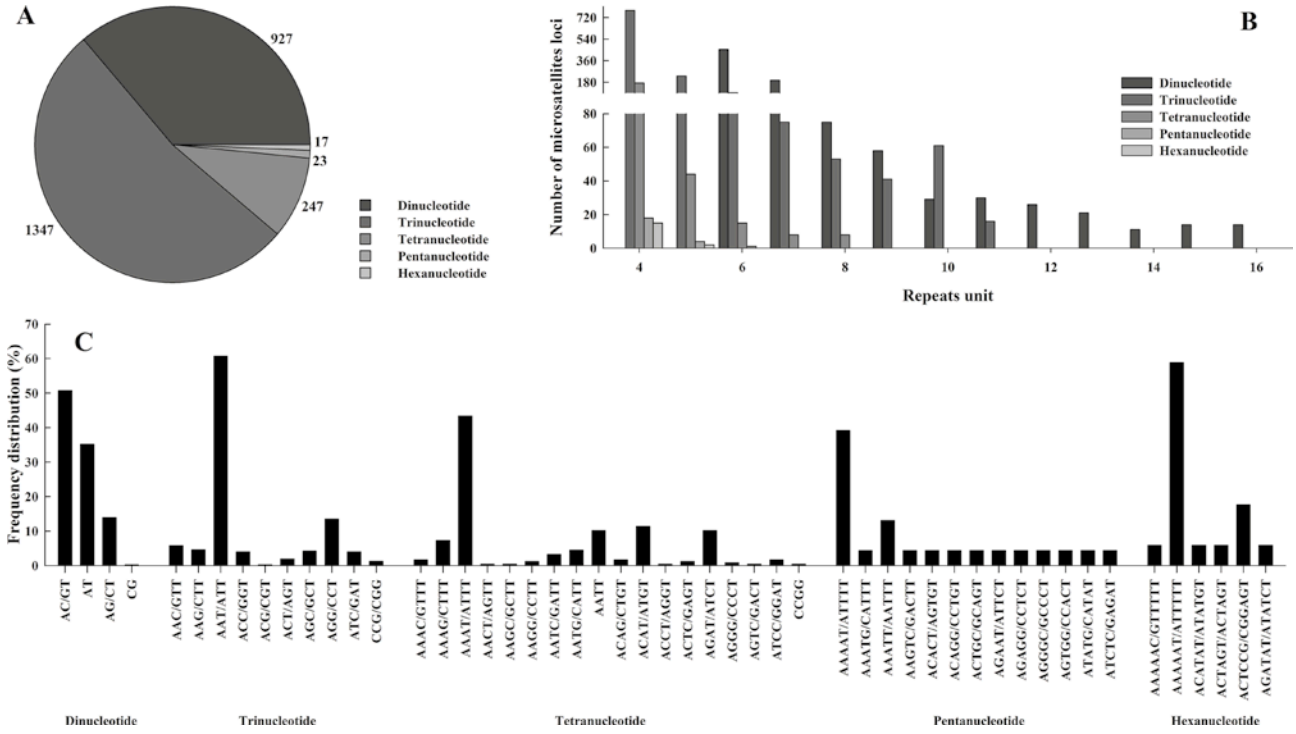

Fig. 1. Characterization of microsatellite loci in *Fejervarya multistriata* genome. (A) distribution of different repeat motif types of microsatellite loci; (B) number of different repeat motifs; (C) frequency distribution of 5 different repeat types based on different motif types. Number represent number of sequence.

respectively. There were 10 trinucleotide motif types, and the main types were AAT/ATT (60.65%) and AGG/CCT (13.51%) (Fig. 1C), respectively. There were 18 tetranucleotide motif types, and the main types were AAAT/ATTT (43.32%), ACAT/ATGT (11.34%), AATT (10.12%) and AGAT/ATCT (10.12%) (Fig. 1C), respectively. There were 13 pentanucleotide motif types, and the main types were AAAAT/ATTTT (39.13%) and AAATT/AATTT (13.04%) (Fig. 1C), respectively. There were 6 hexanucleotide motif types, and the main types were AAAAAT/ATTTTT (58.82%) and ACTCCG/CGGAGT (17.65%) (Fig. 1C), respectively.

The frequency of different repeat types of microsatellites in *F. multistriata* was different from *Xenopus tropicalis* (Xu et al., 2008), *Odorrana narina*, *Hoplobatrachus tigerinus*, and *Buergeria japonica* (Igawa et al., 2015). The dominant repeat type was trinucleotide in *F. multistriata*, but dinucleotide in the last 4 anuran species (Xu et al., 2008; Igawa et al., 2015). Such results may be related to the different next-generation sequencing platforms used in constructing sequence library. Since the library of microsatellite sequences was constructed by Illunima HiSeq platform for *F. multistriata*, however, by Ion Torrent PGM platform for *O. narina*, *H. tigerinus* and *B. japonica* (Igawa et al., 2015). In addition, the results suggested that the frequency of the repeat type changed randomly for each species and was species-specific. The frequencies decreased when the repeat unit length in each repeat type motif of *F. multistriata* increased (Fig. 1B), indicating that a relatively short repeat unit of microsatellites might be a main component in the genome of *F. multistriata.*

Our finding suggested that the ratio of the repeat motifs with an AT content of approximately 56% in our *F. multistriata* was similar to other reported anuran species (e.g., *X. tropicalis*, Xu et al., 2008; *O. narina*, *H. tigerinus* and *B. japonica*, Igawa et al., 2015), suggesting that the AT content could be an important repeat unit in anurans. The dominant repeat motif in the trinucleotide type of *F. multistriata*, (AAT/ATT repeat) was similar to the other four reported anuran species (Xu et al., 2008; Igawa et al., 2015). However, other types of repeat motif were different among these species. For example, *F. multistriata* had a higher frequency in AC/GT and AAAT/ATTT repeats, but *O. narina*, *H. tigerinus*, *B. japonica* and *X. tropicalis* in AT and AGAT/ATCT repeats (Xu et al., 2008; Igawa et al., 2015). These results implied that the accumulation rates of repeat motifs were maintained in modern anurans, but skewed in a common ancestor (Igawa et al., 2015).

Table 1. Characterization of 18 microsatellite DNA markers developed for *F. multistriata*. Size range: size range of fragment; bp: base pair; T_a,: annealing temperature of primer pairs; N_a: number of alleles; H_o: observed heterozygosity; H_E: expected heterozygosity; HWE: Hardy-Weinberg equilibrium; PIC: polymorphic information content; bold: significant deviation from HWE after Bonferroni correction ($P < 0.05$).

Locus (GenBank #)	Primer sequences (5'-3')	Repeat motif	T_a (°C)	Size range (bp)	N_A	H_o	H_E	P_{HWE}	PIC
FMA102 MG744293	F: GCACTGTAGAGCACTGGATTC R: GAGCGTCATAGGGGTCAAATAG	(TA)16	53	129-219	13	0.4792	0.7592	1.000	0.72
FMA349 MG744294	F: CACTCATGTTATCACTCTACTCTC R: CCTCCTACCTCTTGTACTAAATTG	(TAT)11	53	156-237	11	0.3333	0.8145	1.000	0.782
FMA117 MG744295	F: ACTTGAGTCTATTCTATTCTGCTG R: ACTGCTGCTCTGATCTCTATG	(ATA)7	53	149-158	4	0.4167	0.5893	0.996	0.495
FMA402 MG744296	F: AGACATTACCTTAAAGCCATAGTG R: CTTCTGACATGACCTGTTCTTC	(AGAT)6	53	189-201	4	0.4583	0.6090	0.975	0.538
FMA041 MG744297	F: CCAGGAGGATTCTAGTGACAG R: ATGAAGGCAAGAGCAATGTAC	(AT)12	53	152-192	12	0.3958	0.8169	1.000	0.789
FMA466 MG744298	F: GGTGCCACTGTCTTAACTATCC R: AGTCCAATCAAGTCCAATTCAAAC	(TTTTTA)4	53	199-205	2	0.3125	0.4086	0.975	0.323
FMA294 MG744299	F: GTCCTCCTACCTCTTGTACTG R: CGAATGAGAACCTTCACAGAC	(ATA)10	53	187-238	6	0.9375	0.6515	**< 0.01**	0.586
FMA302 MG744300	F: TCCGACCTCTTGAAACTGTATTG R: AGGATCACCACTAGGAGCATC	(ATA)10	53	205-259	13	0.6042	0.8680	1.000	0.845
FMA355 MG744301	F: TATGACCACAGTCTAGCATCC R: CTCCAGTAGTTATCACCTTCTTG	(AAAT)5	58	158	1	–	–	na	0
FMA188 MG744302	F: CCTCTTGTGTTGGTGTATTTCTG R: TTATGCTTGTGTTCTGGTCATTC	(AAT)8	63	209-215	3	0.6042	0.5340	0.170	0.434
FMA231 MG744303	F: GCTGCTGCATGATAGTGTCTC R: TGATGTCTGATGGTCGTCCTG	(TAT)8	53	155-185	10	0.7917	0.8353	0.992	0.805
FMA072 MG744304	F: TGCAGTAGACATCGGAGTTG R: GCCTCTCTCATCTTATTAAGTGG	(TA)13	53	209-237	10	0.6875	0.7934	0.998	0.757
FMA140 MG744305	F: TTCATTGTGCCAAGTGTAACG R: TAACAAAGAGGTCATCACTAATCC	(ATT)7	63	153-165	3	0.1875	0.2254	0.927	0.206
FMA403 MG744306	F: GCGTGGATCGTTATTGAAGTG R: GGTGACCTAATGTGAAATTCCTG	(ATTT)6	63	193-197	2	0.2708	0.2945	0.858	0.249
FMA116 MG744307	F: CTCCCTAACTATTGTAAAGCACTG R: ATTATAGATGGAAGCAACAGGAAC	(ATA)7	55	165-195	8	0.5208	0.7825	1.000	0.743
FMA399 MG744308	F: TTCAGGCTACAGGCATTACAG R: ATAAGGGTGTTCTGCTAAATCAAG	(AATA)6	55	168-228	4	0.2708	0.4476	1.000	0.416
FMA269 MG744309	F: AATGCTTGCAGAACTATTCACAC R: TACGGCGGTCCTAAGATGG	(TAT)9	55	177-192	6	0.2500	0.6732	1.000	0.616
FMA139 MG744310	F: GATTGATGGATTGATGATGGACTG R: AATGTTCAAGATGGACGAATTACC	(ATG)7	55	177-189	5	0.6042	0.6535	0.813	0.584

Characterization of microsatellite loci

Twenty-six primer pairs were used to successfully amplify genomic DNA of *F. multistriata* from LST population. Of the 26 pairs, 18 pairs generated target bands, and the other 8 pairs generated non-target bands. Finally, a total of 18 primer pairs were characterized, of which 17 were polymorphic. The genomic sequences containing a microsatellite locus, which were used to design these primers, were deposited in GenBank (accession number: MG744293–MG744310). The information of primer sequences, repeat motifs, T_a, N_A, PIC and heterozygosity for each locus were shown in Table 1. The N_a, PIC and heterozygosities (H_O and H_E) ranged from 1 to 11 (4.815 ± 2. 699, mean ± SD), 0 to 0.845 (0.549 ± 0.240, mean ± SD), 0 to 1.0 (0.452 ± 0.246, mean ± SD), 0 to 0.871 (0.571 ± 0.237, mean ± SD) within each population, respectively. No significant linkage disequilibrium was observed after Bonferroni correction for multiple tests (all $P > 0.05$). Of the 18 loci, one (FMA294) deviated significantly from HWE testing each population separately and combined (all $P < 0.05$; Table 1), which indicated that null alleles may be present in this locus (Song et al., 2017). The locus was assessed to contain moderately high

polymorphism degree when the value of PIC was larger than 0.5 (Song et al., 2017). Overall, 11 loci had high polymorphism degree (PIC > 0.5), 6 loci had low polymorphism degree (PIC < 0.5). These informative microsatellite loci will be applicable for conservation genetic studies of *F. multistriata* across varying scales from inter-individual to inter-population.

ACKNOWLEDGMENTS

We thank Yao-Fei Yu, Zhi-Qiang Chen for their help in field work. This study was funded by National Science Foundation of China (31500308), Zhejiang Provincial Natural Science Foundation of China (LQ16C040001), Key Research Projects of Lishui City (SH2017001), Zhejiang Science and Technology Innovation Program for College Students (2018R434003, 2019R434006).

REFERENCES

Abe, H., Hayano, A., Inoue-Murayama, M. (2012): Forensic species identification of large macaws using DNA barcodes and microsatellite profiles. Mol. Biol. Rep. **39**: 693-699.

AmphibiaChina (2019): The database of Chinese amphibians. Kunming Institute of Zoology (CAS), Kunming, Yunnan, China. Available from: http://www.amphibiachina.org/ [*accessed on 1 March 2019*]

Chang, Y.C., Li, S.H., Lin, H.Y., Chen, S.L., Chang, M.H. (2016): Development of 22 polymorphic microsatellite markers for Taipei grass frogs (*Hylarana taipehensis*). Amphibia-Reptilia **37**: 117-120.

Chen, Y.D., Li R., Wei, L., Lin, Z.H., Li, Y.D., Ding, G.H. (2018): Development and characterization of 21 microsatellite loci for *Fejervarya multistriata* (Anura: Dicroglossidae) using genome-wide data. Microsatellite records for volume 10, issue 1. Conserv. Genet. Resour. **10**: 127-140.

Ellegren, H. (2004): Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. **5**: 435-445.

Goudet, J. (1995): FSTAT (Version 1.2): a computer program to calculate F-statistics. J. Hered. **86**: 485-486.

Guo, K., Ding, G.H., Sun, Y.Y., Lu, H.L. (2015): Isolation and characterization of microsatellite loci in the northern grass lizard, *Takydromus septentrionalis* (Squamata: Lacertidae). Conserv. Genet. Resour. **7**: 583-584.

Igawa, T., Nozawa, M., Nagaoka, M., Komaki, S., Oumi, S., Fujii, T., Sumida, M. (2015): Microsatellite marker development by multiplex Ion Torrent PGM sequencing: a case study of the endangered *Odorrana narina* complex of frogs. J. Hered. **106**: 131-137.

Li, B., Zhang, W., Shu, X.X., Pei, E.L., Yuan, X., Sun, Y.J., Wang, T.H., Wang, Z.H. (2016): The impacts of urbanization on the distribution and body condition of the rice-paddy frog (*Fejervarya multistriata*) and gold-striped pond frog (*Pelophylax plancyi*) in Shanghai, China. Asian Herpetol. Res. **7**: 200-209.

Lü, Z., Li, H., Liu, L., Cui, W., Hu, X., Wang, C. (2013): Rapid development of microsatellite markers from the large yellow croaker (*Pseudosciaena crocea*) using next generation DNA sequencing technology. Biochem. Syst. Ecol. **51**: 314-319.

Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton J.M. (1998): Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**: 639-655.

Othman, M.S., Khonsue, W., Kitana, J., Thirakhupt, K., Robson, M. G., Kitana N. (2009): Cadmium accumulation in two populations of rice frogs (*Fejervarya limnocharis*) naturally exposed to different environmental cadmium levels. B. Environ. Contam. Tox. **83**: 703-707.

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

Selkoe, K.A., Toonen, R.J. (2006): Microsatellite for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. **9**: 615-629.

Shao, W.W., Lin, Z.H., Fan, X.L., Ding, G.H., Wei L. (2017): Characterization of 25 microsatellite loci for the Guenther's frog *Hylarana guentheri*. Conserv. Genet. Resour. **10**: 31-33.

Song, W., Zhu, D.M., Lv, Y.F., Wang, W.M. (2017): Isolation and characterization of 37 polymorphic microsatellite loci of *Megalobrama hoffmanni* by next-generation sequencing technology and cross-species amplification in related species. J. Genet. **96**: 39-45.

Soulard, L., Mournet, P., Guitton, B., Chaïr, H. (2017): Construction of two genetic linkage maps of taro using single nucleotide polymorphism and microsatellite markers. Mol. Breeding. **37**: 37.

Sultana, N., Igawa, T., Nozawa, M., Islam, M.M., Hasan, M., Alam, M.S., Khan, M.M., Sumida, M. (2014): Development and characterization of 27 new microsatellite markers for the Indian bullfrog *Hoplobatrachus tigerinus* and its congeneric species. Genes Genet. Syst. **89**: 137-141.

Wambulwa, M.C., Meegahakumbura, M.K., Chalo, R., Kamunya, S., Muchugi, A., Xu, J.C., Liu, J., Li, D.Z., Gao, L. M. (2016): Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. Tree Genet. Genome. **12**: 11.

Wei, L., Li, L.M., Chen, J.P., Shao, W.W., Lin, Z.H. (2015): Characterization of 14 microsatellite loci for the pan-oriental narrow-mouthed toad, *Microhyla ornata* (Amphibia: Microhylidae). Conserv. Genet. Resour. **7**: 33-35.

Xu, Z.K., Gutierrez, L., Hitchens, M., Scherer, S., Sater, A. K.,Wells, D.E. (2008): Distribution of Polymorphic and Non-Polymorphic Microsatellite Repeats in *Xenopus tropicalis*. Bioinform. Biol. In. **2**: 157-169.

Yang, T., Bao, S.Y., Ford, R., Jia, T.J., Guan, J.P., He, Y.H., Sun, X.L., Jiang, J.Y., Hao, J.J., Zhang, X.Y., Zong, X.X. (2012): High-throughput novel microsatellite marker of faba bean via next generation sequencing. BMC Genomics **13**: 602.

Zhang, L.N., Peng, J., Li, X. J., Liu, Y.L., Cui, C. J., Wu, H., Wu, R.N., Tian, P.P., Li, Y. (2013): Development of 27 trinucleotide microsatellite markers for *Saccharina japonica* using next generation sequencing technology. Conserv. Genet. Resour. **6**: 341-344.

Zhong, J., Liu, Z.Q.,Wang, Y.Q. (2008): Phylogeography of the rice frog, *Fejervarya multistriata* (Anura: Ranidae), from China based on mtDNA D-loop sequences. Zool. Sci. 25: 811-882.