Turk. J. Fish.& Aquat. Sci. 21(6), 299-308
http://doi.org/10.4194/1303-2712-v21_6_05

RESEARCH PAPER

Development and Characterization of 68 Microsatellite Markers of Black Amur Bream *Megalobrama terminalis* by Next-Generation Sequencing

Kai Liu[1], Xiao-yu Feng[1], Heng-jia Ma[1], Nan Xie[2,*]

[1]Hangzhou Academy of Agricultural Sciences, Institute of Fishery Science, Hangzhou, Zhejiang, China

How to cite

Liu, K., Feng, X., Ma, H., Xie, N. (2021). Development and Characterization of 68 Microsatellite Markers of Black Amur Bream Megalobrama terminalis by Next-Generation Sequencing. *Turkish Journal of Fisheries and Aquatic Sciences*, *21*, 299-308. http://doi.org/10.4194/1303-2712-v21_6_05

Article History
Received 16 November 2020
Accepted 02 March 2021
First Online 04 March 2021

Corresponding Author
Tel.: +86057187647786
E-mail: n.xie@163.com

Keywords
Megalobrama terminalis
Next-generation sequencing
Microsatellite
Polymorphism

Abstract

This study developed and characterized 68 novel polymorphic microsatellite markers from black Amur bream *Megalobrama terminalis* by next-generation sequencing. Variability was tested on 36 individuals collected from Qiantang River, Zhejiang Province, China. The number of alleles per locus ranged from 2 to 13. Observed heterozygosity ranged from 0.028 to 0.944, whereas the expected heterozygosity ranged from 0.028 to 0.887. Polymorphism Information Content ranged from 0.027 to 0.862. Moreover, 53 microsatellites were in agreement with Hardy–Weinberg equilibrium. Twenty-eight pairwise tests in 33 microsatellite loci indicated linkage disequilibrium. These microsatellites are a valuable tool for further genetics studies of this species.

Introduction

Black Amur bream *Megalobrama terminalis* (Richardson, 1846), a Cyprinidae family member, is a benthopelagic freshwater omnivorous fish (Froese & Pauly, 2019). It is endemic to Asia and widely distributed from the Amur basin between the Russian Far East and Northeastern China to Southern China's rivers (Froese & Pauly, 2019). *Megalobrama terminalis* is a delicious and nutritious fish. Therefore, it is regarded as a high-quality fish and deeply favored by consumers. However, in recent years, this fish's wild resources have declined due to overfishing, water pollution, and growing market demand (Hu & Shi, 2020). At present, the Qiantang River in East China has a certain amount of resources due to developing a protective plan for local *M. terminalis* germplasm resources and has the only national original breeding farm of *M. terminalis* in China (Hu & Shi, 2020). Nevertheless, the conservation of the natural populations of this species is an urgent need.

Genetic data for this organism is scarce. Scientific studies are also necessary for the conservation of this species. Researches based on molecular markers are an essential method for estimating polymorphism and structures in wild populations. Microsatellites are often referred to as simple sequence repeats (SSRs). SSRs are highly polymorphic codominant markers and are suitable for genetic diversity assessments in fish (Chistiakov *et al.*, 2006; Ma *et al.*, 2011; Liu *et al.*, 2011; Song *et al.*, 2016; Song *et al.*, 2017). Although twenty-four microsatellite loci excepted from markers developed by us of *M. terminalis* have been registered in GenBank, relevant information on these markers, such as primer sequences and level of polymorphisms, has not yet been published. Thus, the utility of these microsatellite loci as genetic markers remains uncertain.

Published by Central Fisheries Research Institute (SUMAE) Trabzon, Turkey in cooperation with Japan International Cooperation Agency (JICA), Japan

Besides, most of the microsatellite loci registered in GenBank are dinucleotide or mononucleotide repeats. The polymorphism of mono-SSRs is challenging to interpret (Lopez *et al.*, 2015), di-SSRs are more challenging to score accurately because of substantial strand slippage during PCR (Weber *et al.*, 2001), same as tri-SSRs. However, tetra-SSRs reduced strand slippage during PCR compared to dinucleotide repeats (Ghebranious *et al.*, 2003). Moreover, penta-SSRs may have lower stutter proportions than tetranucleotides (De Barba *et al.*, 2017). Meanwhile, a few microsatellite markers developed from *M. amblycephala*, *M. hoffmanni*, and *M. pellegrini* are typically used on *M. terminalis* for genetic diversity evaluation (Zhang *et al.*, 2014; Du *et al.*, 2016; Song *et al.*, 2016; Song *et al.*, 2017). Zhang *et al.* (2014) reported 18 out of 60 polymorphic microsatellite markers from the genomic DNA and transcriptome of *M. amblycephala* could be successfully amplified on *M. terminalis*. Du *et al.* (2016) found that 13 out of 30 microsatellite markers from the *M. amblycephala* transcriptome database could be successfully amplified on *M. terminalis*, and 11 out of 13 loci are polymorphic. Five out of 29 polymorphic microsatellite markers from the genomic DNA of *M. pellegrini* could be successfully amplified on *M. terminalis* (Song *et al*., 2016). Thirty out of 37 microsatellite markers from the genomic DNA of *M. hoffmanni* could be successfully amplified on *M. terminalis*, but only 4 out of 30 loci are polymorphic (Song *et al.*, 2017). In addition, most of the microsatellite markers identified (Zhang *et al.*, 2014; Du *et al.*, 2016) are dinucleotide repeats, which are more difficult to score specifically because of significant strand slippage during the PCR (Weber *et al*., 2001). Thence, in this study, microsatellite loci were targeted to tetra-, penta-, or hexa-SSRs.

Compared with traditional microsatellite marker development methods, next-generation sequencing (NGS) is more cost-efficient (Zheng *et al.*, 2013; Liu *et al.*, 2017). Microsatellite markers derived from expressed sequence tags (ESTs) have achieved high efficiency in gene mapping by EST-simple sequence repeats (EST-SSRs). EST-SSRs are associated with the recognized feature genes and a useful tool for studying a fish population's genetic structure (Gao *et al.*, 2012; Hasselman *et al.*, 2013). Twenty-four microsatellite loci excepted from markers developed by us of *M. terminalis* have been registered in GenBank. However, molecular markers for *M. terminalis* are still not enough, which cannot fully evaluate the germplasm genetics and molecular-assisted breeding system of this species. For example, Bouza *et al.* (2012) constructed a consensus gene-enriched genetic map of the turbot based on 463 SNP and microsatellite markers in nine reference families, and Feng *et al.* (2018) used 7,820 SNPs and 295 SSRs to construct the common carp's high-density linkage map. Therefore, more molecular markers for this fish urgently need to be developed.

Materials and Methods

Sample Collection and Genomic DNA Extraction

M. terminalis utilized in this study were randomly harvested from the national original breeding farm of black Amur bream from Qiantang River, Zhejiang province, China. Fin clips samples were used, obtained, and stored in anhydrous alcohol at −20°C before DNA was extracted. Genomic DNA derived by M. terminalis was achieved following a standard phenol-chloroform extraction method (Green & Sambrook, 2012).

RNA Extraction and Sequencing

Transcriptomes of liver tissue were sequenced to access many diverse transcripts because the liver is a highly complex organ with a complex transcriptome (Shackel et al., 2002; Shackel et al., 2006). The livers of 6 M. terminalis were collected and mixed, used as a sequence sample to build a cDNA library. The sample tissues were frozen in liquid nitrogen and stored at −80°C until use. The PCR products were sent to Personal Gene Technology CO., Ltd (Shanghai, China) for HiSeq™ 2000 sequencing.

EST-SSR Detection and Primer Development

Microsatellites within the unigenes assembly were detected using a Perl script MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html). The SSR loci were considered to target microsatellites with at least five tandem repeats for tetra-, penta -, and hexa-SSRs. EST-SSRs primers were designed using Primer3 ver. 2.3.6 (http://sourceforge.net/projects/primer3) under the following criteria, primers' length is approximately 20 bp. The melting temperature was around 60°C.

PCR Conditions and Amplification of Microsatellites

Genomic DNAs from 3 M. terminalis individuals were used to initially screening primers that TouchDown PCR can amplify. According to the manufacturer's instructions, PCR amplifications were conducted using AmpliTaq Gold™ Fast PCR Master Mix (Applied Biosystems, Foster City, CA, USA). PCR amplifications were conducted at the following conditions: initial denaturation for 3 min at 94°C, followed by ten cycles for 30 s at 94°C, 30 s at the annealing temperature, and 20 s at 72°C; the initial annealing temperature was 65°C, which reduced by 1°C in each cycle, followed by 20 cycles for 30 s at 94°C, 30 s at 55°C, and 20 s at 72°C; that was followed by a final extension at 72°C for 5 min. Amplification products were analyzed by 3% agarose gel electrophoresis. The loci which cannot be amplified in all samples were excluded from further testing. All primers showed stable PCR results were labeled with a fluorescent dye (FAM) on each forward primer.

Genotyping

Microsatellite loci were characterized by 36 individuals of *M. terminalis* randomly harvested from the national original breeding farm of black Amur bream from Qiantang River, Zhejiang province, China. PCR amplifications were conducted as same as the procedure described above. PCR amplifications were also carried out with the related parameters described above, while the final extension for PCR was conducted at 72°C for 30 min. The PCR products were sequenced on an ABI3730xl automated sequencer (Applied Biosystems, Foster City, CA, USA). The fragment size of alleles was determined against the standard size of Liz-500 using GeneMapper software (Applied Biosystems, Foster City, CA, USA).

Microsatellite Data Analysis

The program Cervus ver. 3.0.7 assessed the number of alleles per locus, observed heterozygosity, expected heterozygosity, and the polymorphism information content (*PIC*; Kalinowski *et al.*, 2007). The program PopGene ver. 1.32 assessed confirmation of Hardy–Weinberg equilibrium with the likelihood ratio test (Yeh *et al.*, 1999). The program Arlequin ver. 3.5.2.2 carried out pairwise tests for linkage disequilibrium (Excoffier & Lischer, 2010), performed using a likelihood ratio test whose empirical distribution was obtained by a permutation procedure (Slatkin & Excoffier, 1996). Meanwhile, Bonferroni corrections were used to evaluate the significance (Rice, 1989). The program Micro-Checker ver. 2.2.3 (Van Oosterhout *et al.*, 2004) was used to calculate the null allele frequency.

Results and Discussion

The Gene Ontology (GO) is a controlled vocabulary composed of >38 000 precise defined phrases called GO terms that describe the molecular actions of gene products, the biological processes in which those actions occur, and the cellular locations where they are present (Balakrishnan *et al.*, 2013). GO annotation is the statement of a connection between a type of gene product and the types designated by terms in an ontology (Hill *et al.*, 2008). In this study, 68 microsatellite-containing unigenes were classified by their biological process, molecular function, and cellular component using the Blast2GO program (https://www.blast2go.com/) and visualized by the WEGO program (http://wego.genomics.org.cn/). Among the 68 unigenes, 20 were successfully mapped with GO annotations and classified into three ontologies that contained 23 GO terms (Figure 1). At the cellular GO level, there were nine total GO terms, corresponding to 3 unigenes in the cell part (GO: 0044464), three unigenes in the cell (GO: 0005623), and three unigenes in the membrane (GO: 0016020); Regarding the

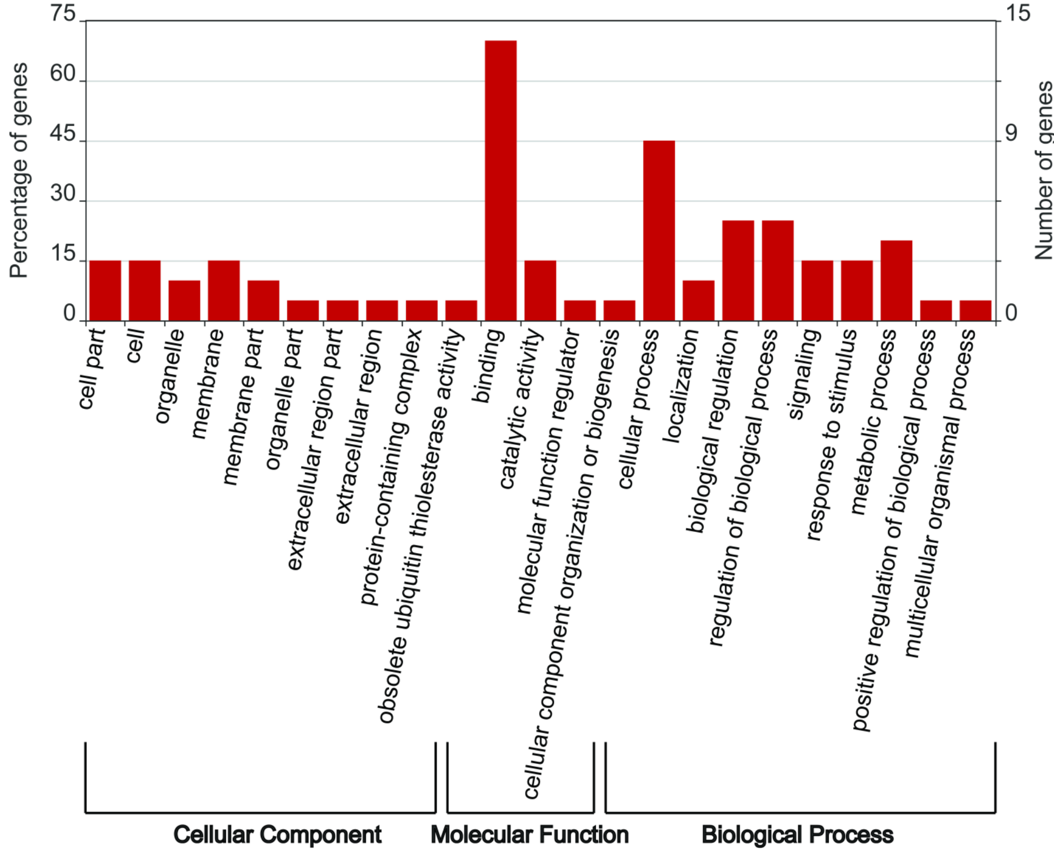

Figure 1. Gene ontology classification of 68 unigenes containing microsatellites in *Megalobrama terminalis*

molecular function ontology, four total GO terms were assigned, and the primary functions were binding functions (GO: 0005488) with 14 unigenes, and catalytic activity (GO: 0003824) with three unigenes. In the biological process category, 10 GO terms were assigned; most of the unigenes were involved in the cellular process (GO: 0009987), biological regulation (GO: 0065007), and regulation of biological process (GO: 0050789).

Sixty-eight unigenes produced by RNA-seq data contained 72 microsatellites with at least five tandem repeats for tetra-, penta-, and hexa-SSRs. There are four unigenes, and each unigenes contains two microsatellite loci. Four loci of the four unigenes were not capable of designing primers. In the 72 microsatellites, Twenty-four motifs were obtained, of which the most frequent was AAAC/GTTT (11, 15.28%), followed by AAAG/CTTT (10, 13.89%), AGAT/ATCT (10, 13.89%), AAAT/ATTT (6, 8.33%), AATC/ATTG (6, 8.33%), AAAAG/CTTTT (4, 4.17%) and AAAAT/ATTTT (4, 4.17%) (Figure 2). Detailed analysis showed that tetra-SSRs were the most frequent (79.17%), followed by penta-SSRs (15.28%) and hexa-SSRs (5.55%). SSRs with five tandem repeats (27, 37.50%) were the most common, followed by six tandem repeats (18, 25.00%) and seven tandem repeats (9, 12.50%).

All 68 microsatellite loci had a polymorphic pattern (Table 1). The number of alleles per locus ranged from 2 (Mt01650, Mt02653, Mt03365, Mt05046) to 13 (Mt01191, Mt07858). The observed heterozygosity ranged from 0.028 (Mt01650, Mt02088) to 0.944 (Mt06235), whereas the expected heterozygosity ranged from 0.028 (Mt01650) to 0.887 (Mt01210). *PIC* ranged from 0.027 (Mt01650) to 0.862 (Mt01210) with 42 out of which being highly informative (*PIC* > 0.5) and 20 moderately informative (0.25 < *PIC* < 0.5) (Botstein *et al.*, 1980). Analysis with the program Micro-Checker showed, in a pool of sampled individuals, low-to-moderate frequencies of null alleles in these loci, which is not surprising. Null alleles are commonly observed in various species (Zhao *et al.*, 2011; Yu *et al.*, 2019). Null frequencies below 0.2 are acceptable in most microsatellite datasets (Dakin & Avise, 2004). In 14 out of 68 microsatellite loci, this estimate was higher than 0.2. According to the analyzed result of Micro-Checker, in nine out of the 14 microsatellite loci, this estimate was higher due to the stutter peak's existence. In the other five out of 14 loci, the most probable reason for this phenomenon being scoring errors may be the heterozygote genotype's loss. Fifteen out of 68 microsatellite loci exhibited significant probabilities of departure from Hardy–Weinberg equilibrium

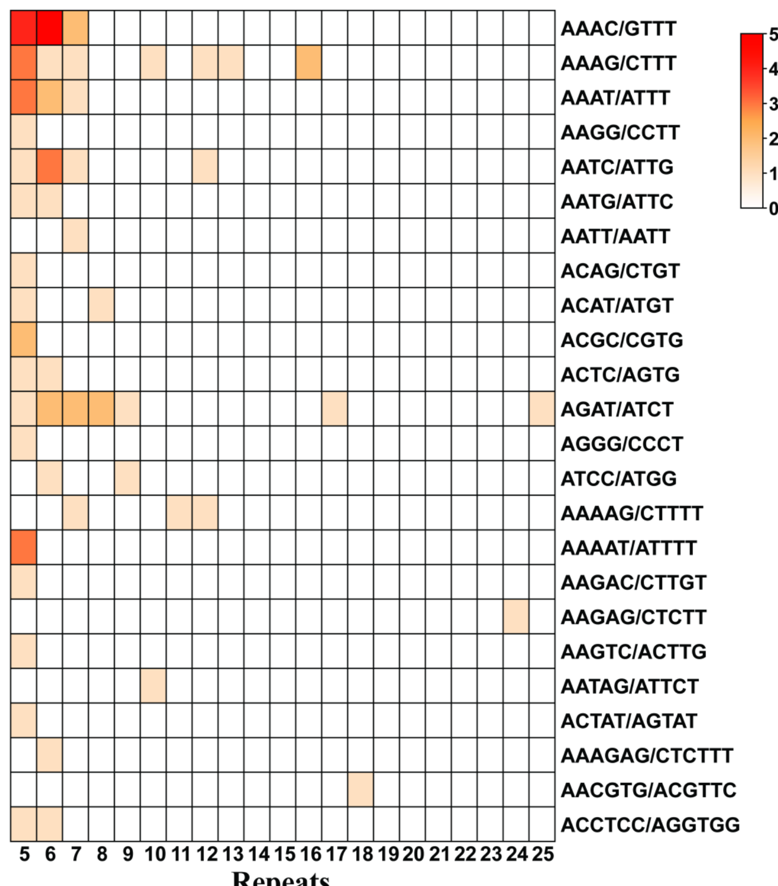

Figure 2. classification of tetra-, penta-, and hexa-SSRs of 68 unigenes in *Megalobrama terminalis* (considering sequence complementary)

Table 1. Characteristics and genetic diversity of 68 microsatellite loci developed in *Megalobrama terminalis*

ID	Repeat motif	Forward Primer(5'-3')	Reverse Primer(5'-3')	Size range	*Na*	*Ho*	*He*	*PIC*	*PHW*	F_{Null}	Accession No.
Mt00453	(TAGTA)5	CCACCAGTGTGAACGCTAAA	TGCTATCCCAAAGAAAACCG	185-195	3	0.361	0.567	0.488	0.0083	0.1269	MT182737
Mt00943	(TGTT)6	ATCCTCATCAGGGACATTGC	AGGTAGAGCAGATGGGCAAA	171-187	4	0.472	0.46	0.391	0.5824	-0.0130	MT182738
Mt00998	(CTTTCT)6	AGGGGGATCGAGAGACAGAT	CACAACACCCTGACAACCAC	226-250	5	0.514	0.636	0.567	0.0396	0.0984	MT182739
Mt01091	(AAAC)6	TGCCATATAGAACAATACATGAAGG	ACCTTATGGACAATGCAGCC	114-130	4	0.306	0.398	0.364	0.0004*	0.0623	MT182740
Mt01146	(TATC)5	AGGAATCCAACCGTCACATC	ATTGTGCTGCTGCATTGAAC	183-207	7	0.629	0.765	0.718	0.6800	0.0909	MT182741
Mt01148	(GAAA)5	CCTAAGGGTGAGGGCTTTTC	TCTTGTCAGGAGTGTGGGTG	130-178	6	0.278	0.647	0.584	0.0000*	0.2200	MT182742
Mt01153	(ATAG)8	TGATCCCCTTTGGTTTTCAG	CACCCCAAGTACCCTAGCAA	263-315	10	0.528	0.751	0.700	0.6176	0.1221	MT182743
Mt01191	(AGAT)7	GTTTCACCAAAACGGGAAAA	GGAGCGAAGTGAAGACCAAG	259-383	13	0.382	0.796	0.763	0.4516	0.3445	MT182744
Mt01210	(TCCA)9	GACTCCTTCACCTGCGTCTC	CTCACTACAGGCCCCCAATA	204-288	12	0.667	0.887	0.862	0.1412	0.1109	MT182745
Mt01650	(TCCACC)5	ACATCAGGTACGGGCTCAAC	CACTCCCACCAGTCAAGGTT	268-274	2	0.028	0.028	0.027	1.0000	-0.0004	MT182746
Mt01770	(AAAG)5	TGGGCTTTTGAGGGTACAAA	ATGCGTTTCTTGGCTTGACT	189-281	7	0.444	0.720	0.665	0.0207	0.1552	MT182747
Mt02032	(ATCA)12	GGCCCAGTGACCTTCAACTA	CCCCTTTAAAATCCTTGGGA	79-115	7	0.686	0.700	0.651	0.5949	0.0030	MT182748
Mt02088	(ATTTT)5	GATGCCGCCAAGAACTAAAA	ATTCCAAGCAAGGCAAAGAG	263-278	4	0.028	0.184	0.176	0.0001*	0.1303	MT182749
Mt02229	(TATT)6(TGTT)6	AAATTGGCGATGAGAACTGG	CAGCTGCTAGTAATTTTAAGCTCCA	212-276	8	0.361	0.783	0.736	0.0001*	0.2320	MT182750
Mt02292	(CTTT)16	TGGTTTATTGCCTGCTTGGT	GAGCTAAAGATCAAAAACATGCTC	213-261	10	0.500	0.646	0.612	0.7871	0.0839	MT182751
Mt02417	(TATT)6(TGTT)6	CGATGAGAACTGGGCAAGA	TGGACCAGAGAGAAATGATGC	235-351	12	0.343	0.719	0.682	0.2466	0.3481	MT182752
Mt02610	(AAGA)6	AATTTAACCAACACTTTATTACCAAGA	TCCAGACCTTCATGTGTCCA	211-223	4	0.194	0.182	0.171	0.9956	-0.0124	MT182753
Mt02653	(GATG)6	GCCATTACTGCTGGATGGTT	ATGAGATATTGGCGTCCGTC	270-278	2	0.472	0.504	0.373	0.7053	0.0162	MT182754
Mt02681	(AAGAC)5	TGGGGTCCATTTAATTCCAA	TGATGGCATGAGAACGGTAA	189-264	4	0.389	0.534	0.415	0.4005	0.0902	MT182755
Mt02770	(TTCA)6	ACATTTGTATCGCTCGGAGG	CTGTCGATGCCAGCTCAATA	248-272	5	0.528	0.464	0.416	0.0659	-0.0485	MT182756
Mt03071	(AAAG)7	GGACAATTTTGATTTCAAAGGC	CCTGTAGATGGCAAGGTGGT	202-258	6	0.389	0.477	0.425	0.5713	0.0556	MT182757
Mt03327	(TGAT)6	CTGCAGTGTTCCTCACATGG	TCATTCACAGTAACTGCCTTTCA	251-259	3	0.472	0.548	0.465	0.0920	0.0443	MT182758
Mt03365	(ATGT)8	TTTTCTGAGGTGGGATGGAC	GATCAGAAAATCCTGCTCCG	221-225	2	0.444	0.501	0.372	0.4936	0.0331	MT182759
Mt03456	(GAAA)13	AGATGCCGAAAGCTTGTGAT	CAAAGATCGGAAGGCTGGTA	239-271	9	0.528	0.772	0.733	0.0695	0.1325	MT182760
Mt03576	(TTTG)5	TTGCCGCCTTAAGTCAAAGT	ATCTAGACGGCATTTCGGTG	177-205	6	0.917	0.711	0.644	0.0111	-0.1268	MT182761
Mt03643	(GAGT)5	GGCGTGTCCTGAGTTTTGAT	ACGTACATTCAGGGCGTCTC	89-101	4	0.833	0.632	0.578	0.0003*	-0.1295	MT182762
Mt03873	(TGACT)5	GGCTGCTTTGACACAATCTG	ATCGCATATCGTTACAGCCC	197-262	8	0.361	0.642	0.589	0.1485	0.1667	MT182763
Mt04004	(AAAC)5	GCTTGCTTATGTCCACACCA	TGAATTTTGTGAGGGGTCAA	251-267	3	0.222	0.286	0.249	0.5038	0.0466	MT182764
Mt04221	(TTTTC)12	TGGCTGCAGATGAAATTAAGG	GGTTTGCACAAGTTGGTGTG	218-278	10	0.306	0.838	0.805	0.0002*	0.2852	MT182765
Mt04243	(AAAT)5	TACAATGCGATAGTGCTGCC	GCCGCCAGATATTCTTCAAA	213-225	4	0.306	0.435	0.396	0.0122	0.0861	MT182766
Mt04398	(TTTTC)7	AGATTCTGGCTGCAGATGAAA	GGTTTGCACAAGTTGGTGTG	224-289	8	0.194	0.747	0.698	0.0000*	0.3123	MT182767
Mt04429	(TAGA)6	GCACTGGCAGAAATCCAAGT	ATTCTGAGCACCAGATGGGT	179-195	5	0.694	0.684	0.616	0.0204	-0.0120	MT182768
Mt04680	(CAAA)7	GCGTCAGAAGAGAAGGACTGA	TCAGACGCTTGTGAGGAATG	207-275	8	0.750	0.789	0.744	0.0638	0.0156	MT182769
Mt04878	(ACGC)5	TTACTCCAGAGCACACGCAC	GGCAGTCATGCACTTTCTGA	100-108	3	0.333	0.502	0.397	0.0041	0.1084	MT182770
Mt04992	(TAGA)6	GCAGCCTTGGTGAGCATAAG	GCAGTGGATACATCAGACAGGA	188-220	6	0.417	0.668	0.621	0.0000*	0.1458	MT182771
Mt05046	(TCAA)5	AGCTGTAGCCGAGATCCTCA	CTCTCACATGTTTGTGCGGT	192-200	2	0.500	0.505	0.374	0.9474	-0.0010	MT182772
Mt05103	(CCTC)5	TGTTATCGGTTCTCCTCAGCA	CTTTGAACGTGTCTGCGGTA	202-246	6	0.611	0.696	0.623	0.8766	0.0444	MT182773
Mt05108	(TTTC)5	GCAGGTATGCACCCCTGATA	CTCAATAAGCAGCCCAAAGG	187-203	5	0.444	0.554	0.449	0.1837	0.0661	MT182774

Table 1. *Continued*

ID	Repeat motif	Forward Primer(5'-3')	Reverse Primer(5'-3')	Size range	*Na*	*Ho*	*He*	*PIC*	*PHW*	F_{Null}	Accession No.
Mt05212	(GATT)7	AAACCATCTTCATCAACCGC	CGAGCCAAACTTACCTGTCC	112-132	5	0.722	0.584	0.525	0.0012	-0.0930	MT182775
Mt05488	(GTGA)6	CGAAGACTCGTTCTGGTTGC	GAAAAACTGCCAGGGAAACA	219-243	7	0.639	0.737	0.683	0.7312	0.0512	MT182776
Mt05531	(CTTC)5	ACTACAGTACCTGCCGCTCC	CTGAGCATCATATACAGAGGCA	266-298	7	0.611	0.664	0.631	0.5160	0.0263	MT182777
Mt05575	(TTTC)10	TGGTCCCCAGGACATAATTT	CCTGCTGGTCAACAGAAAGA	254-278	5	0.917	0.777	0.726	0.0753	-0.0850	MT182778
Mt05612	(TATT)7	AAATTGGCGATGAGAACTGG	TGCAAAATGTGATAATTCACGA	248-280	9	0.389	0.805	0.768	0.0013	0.2258	MT182779
Mt05670	(AAAG)12	TTGGTAGTAAACTGCCATTTATTCAG	CCTGTAGATGGCAAGGTGGT	139-175	4	0.361	0.466	0.403	0.4236	0.0674	MT182780
Mt05830	(TAGA)8	GCATTTCCCAAAGAAGAGCA	AGTGTTTATGGCCGTTTTCG	112-152	11	0.806	0.828	0.793	0.3704	0.0062	MT182781
Mt06059	(TTTG)5	GCAGGTCCAAATCTGTCCAT	CGCTTTGACACCACTTTTTG	153-169	5	0.806	0.685	0.616	0.0171	-0.0778	MT182782
Mt06205	(TTCA)5	CTGCAGCGTTGTAATGGAGA	CAAGCTTAGCCCACAGACCT	127-135	3	0.889	0.598	0.517	0.0000*	-0.1881	MT182783
Mt06235	(AGAAG)24	CTTCAACATGAAGCACGCAT	GCAGGAGAGGCAGAAACAAC	174-229	10	0.944	0.806	0.768	0.5499	-0.0836	MT182784
Mt06290	(GTCT)5	CAATCACGCCTCTCTTCTCC	ATGCTCTTCTTGGGACGAAA	232-240	3	0.611	0.579	0.478	0.0152	-0.0253	MT182785
Mt06545	(AAAAT)5	CCTGAAACTACTTGACCGGC	AATTTACAAAAAGCGCCGAG	224-234	3	0.278	0.246	0.219	0.6919	-0.0283	MT182786
Mt07071	(ATTTT)5	AAAGTCTGCCCTAACTATCCTCC	GATACATATGCAGGGTGGGG	217-232	3	0.056	0.055	0.054	0.9987	-0.0011	MT182787
Mt07184	(GTTT)5	TGGCAGGAAGTTGGTTCTTT	GTTGTTGAAGCCCCCAAGTA	225-245	4	0.278	0.446	0.365	0.1043	0.1128	MT182788
Mt07497	(TTTTC)11	TCCTTTGGGGAGAGAGGAGT	TGCAAGGATAGGGGTGCTAT	179-236	10	0.528	0.814	0.776	0.0037	0.1526	MT182789
Mt07649	(TAGA)25	AACACGAGCAGAGCATCAGA	TTTTGGCAGATTGATTTCCAC	171-227	8	0.389	0.806	0.767	0.0004*	0.2260	MT182790
Mt07858	(AGAT)9	TGCAGCCTTTACATGCTAGTG	CAAGCTGCTCCTTCTTGTGA	170-238	13	0.444	0.835	0.805	0.1819	0.2080	MT182791
Mt08470	(TTTA)5	GACCTGGATGCTGAACACCT	GGGGTGACTGAACTAATGTAGCA	210-246	5	0.444	0.723	0.665	0.0000*	0.1566	MT182792
Mt08751	(TTGT)7	TGTCCATCCATACATCCCCT	ATAACACATCGCTTCCCTGG	150-174	4	0.556	0.531	0.413	0.8570	-0.0210	MT182793
Mt08967	(TTAA)7	TTGAGCAGAGTTTCAGTCTGTTTC	TTTCCCTGACCCTGTCAATC	150-238	7	0.278	0.642	0.594	0.0006*	0.2174	MT182794
Mt08993	(TATG)5	TGCCTGTTCTTCAGGTTTCA	GGAAATAATGCACTTGGACACA	140-168	7	0.167	0.423	0.395	0.0077	0.1769	MT182795
Mt09110	(CAAA)6	ATGGCACCCACTTTGACATT	ACCCTGCACATTTTGACACA	221-249	7	0.222	0.685	0.621	0.0001*	0.2706	MT182796
Mt09258	(ACGC)5	TCAGATTCTTGGGCGTTTTC	CCCATTTCTGTGTGCAAATG	241-281	7	0.500	0.790	0.748	0.0008	0.1570	MT182797
Mt09745	(GGTGGA)6	GGCCGTCTTTGGTATGTGTT	ACATCAGGTACGGGCTCAAC	163-223	6	0.361	0.665	0.596	0.0032	0.1780	MT182798
Mt09943	(AAAG)16	CTTTGAAGCTGGTATGGCGT	ATTAGCAATGCCTCCTCCCT	216-256	8	0.250	0.766	0.728	0.0000*	0.2881	MT182799
Mt09962	(AATAG)10	CAAACGACGTATTTGTACTGCG	TGGGTGCGAAGTGTTATTATAGA	130-275	8	0.444	0.523	0.495	0.4547	0.0468	MT182800
Mt10251	(TATC)17	CAAGCAATGCTCTCACAACC	TCTGTGTGGATGCTAGGGTG	188-248	6	0.139	0.301	0.284	0.0311	0.1217	MT182801
Mt10425	(CGTTCA)18	TAACCTGCTGTCGGTGAGTG	AATGCGGGACATTTTCTCAA	185-239	7	0.139	0.741	0.698	0.0000*	0.3418	MT182802
Mt10537	(AATA)5	TGACATCTTCACCCATCCAA	GCCTCCATTTCATATTTCAAGA	233-265	7	0.806	0.693	0.630	0.0329	-0.0724	MT182803
Mt10953	(AGAT)7	CGACATGAGCCTCAATTGTTT	CTTTGGAATAACGGCTTGGA	229-285	10	0.417	0.820	0.786	0.0050	0.2166	MT182804

Note: *Na* indicates the number of alleles per locus, *Ho* indicates observed heterozygosity, *He* indicates expected heterozygosity, *PHW* indicates probabilities derived from Hardy-Weinberg equilibrium,*indicates a significant deviation after Bonferroni correction ($P < 0.0007$), *PIC* indicates polymorphism information content, F_{Null} indicates null allele frequency.

expectations after Bonferroni correction, adjusted critical $P < 0.0007$. The deviations occurred, perhaps resulting from the presence of null alleles or small sample size, or considering the potential for extensive gene flow in this species, the Wahlund effect (Johnson & Black, 1984). Of the 2278 pairwise tests on the 68 microsatellites, 28 pairwise tests indicated linkage disequilibrium, adjusted critical $P < 0.0007$ (Figure 3). The linkage disequilibrium patterns may affect genome-wide association studies' success and the genomic selection and provide key information about demographic history (Yoshida *et al*., 2019). In this study, 28 pairwise tests in 33 microsatellite loci (Mt01148, Mt01210, Mt01770, Mt02088, Mt02229, Mt02292, Mt02610, Mt02681, Mt03071, Mt03327, Mt03456, Mt04221, Mt04398, Mt04680, Mt04878, Mt04992, Mt05575, Mt05612, Mt06059, Mt07184, Mt07497, Mt07649, Mt07858, Mt08470, Mt08751, Mt09110, Mt09745, Mt09943, Mt09962, Mt10251, Mt10425, Mt10537, Mt10953) indicated linkage disequilibrium among loci, which might be due to natural selection.

Using the fast isolation by AFLP of sequences containing repeats (FIASCO) method, we have developed 15 microsatellite markers from the genomic DNA of *M. terminalis* (Liu *et al*., 2020). However, most of the markers developed by the FIASCO method are di-SSRs. Besides, most of the polymorphic SSRs reported by Zhang *et al.* (2014) are di-SSRs, too. Furthermore, 11 out of 30 polymorphic SSRs reported by Du *et al.* (2016) are almost di-SSRs. di-SSRs are more challenging to score accurately because of substantial strand slippage during PCR (Weber *et al.*, 2001). Song *et al*. (2016, 2017) had developed 5 out of 29 and 4 out of 37 polymorphic tetra-, penta-, and hexa-SSRs from the genomic DNA of *M. pellegrini* and *M. hoffmanni,* respectively, that could be successfully amplified on *M. terminalis*. However, only 9 SSRs are insufficient to evaluate germplasm genetics, especially constructing a microsatellites-based linkage map. In contrast, 68 polymorphic microsatellite markers have been developed in this research. These markers significantly increased the number of microsatellite markers of *M. terminalis*.

The SSR markers can be classified into two groups, EST-SSRs and the genomic SSRs (gSSRs). In general, the frequency of EST-SSRs can be confirmed to be decreased because SSRs have a high degree of mutation and can influence gene expression (Vieira *et al.*, 2016). However, some studies indicate a predominance of the tri-SSRs and hexa-SSRs in coding regions, resulting from selection pressure against mutations that alter the reading frame (Zhang *et al*., 2004; Xu *et al*., 2013). In humans, tandem repeats have been documented to be

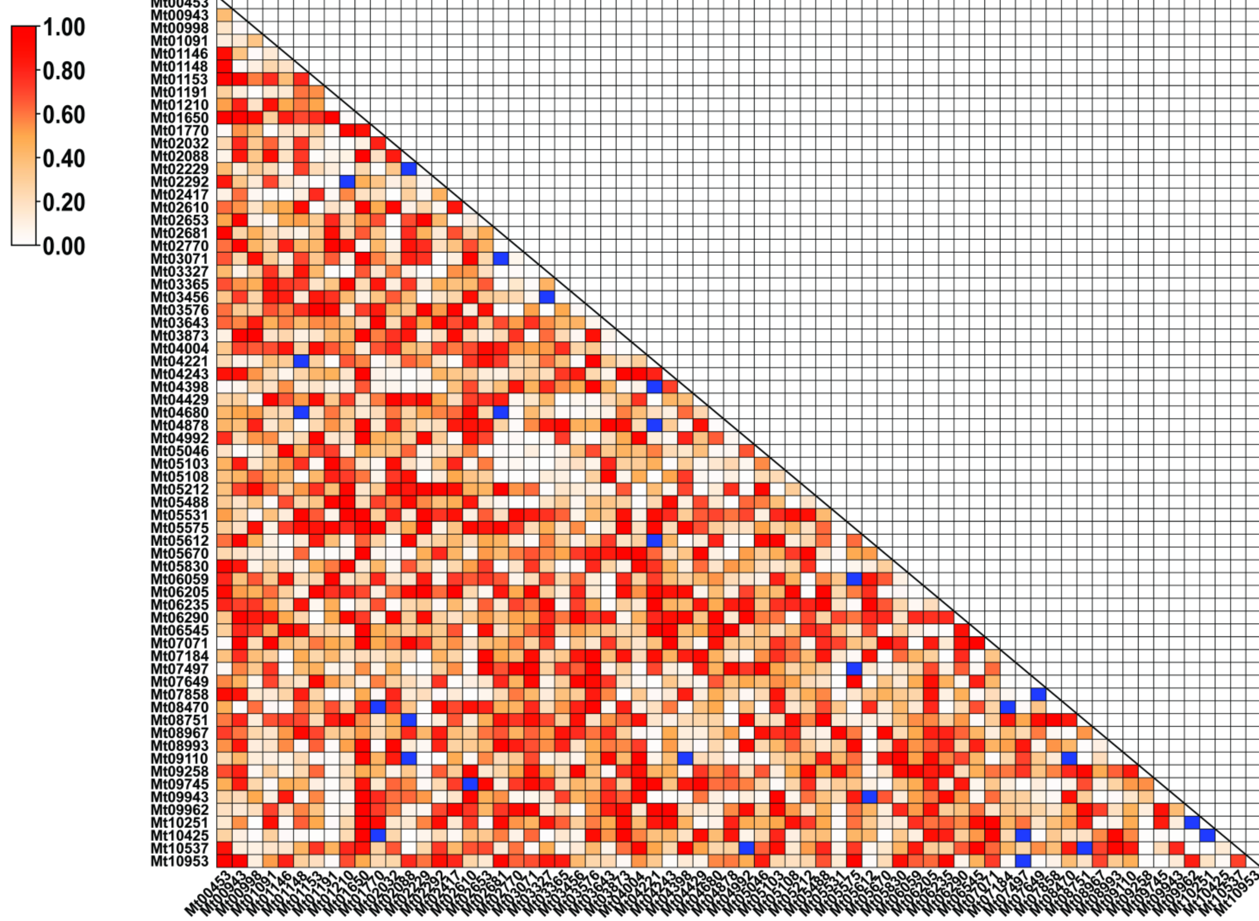

Figure 3. the probabilities from pairwise tests for linkage disequilibrium of 68 microsatellite loci in *Megalobrama terminalis*. Blue squares indicate a significant deviation after Bonferroni correction ($P < 0.0007$)

expected in many proteins, and the mechanisms involved in their genesis may contribute to the rapid evolution of proteins (Vieira *et al.*, 2016). Different SSRs have different uses. gSSRs are resulting in better map coverage, and utilization in comparative mapping and evolution studies is advantageous to EST-SSRs (Parthiban *et al*., 2018). The SSRs in this research are isolated within the unigenes. Therefore, such SSRs are not neutral. EST-SSRs are treated as non-neutral markers but useful for studying adaptive genetic diversity (Ellis & Burke, 2007). Although the developed 68 SSRs are a valuable tool for further genetics studies of this species, they should be treated with caution due to the significant Hardy-Weinberg and linkage disequilibrium tests. New studies with different populations should be employed to evaluate the potential of the developed SSRs. It is expected that the reduction of the NGS cost will increase the diffusion of this approach for non-model organisms. NGS will uncover DNA polymorphisms at an unprecedented scale by making available extensive data on both gSSRs and EST-SSRs (Manco *et al*., 2020).

Ethical Statement

The guidelines established by the Administration of Affairs Concerning Animal Experimentation state that approval from the Science and Technology Bureau of China and the Department of Wildlife Administration is not necessary when the fish in question are neither rare nor near extinction (first- or second-class state protection level). Therefore, approval was not required for the experiments conducted in this study.

Funding Information

This study was supported by the Earmarked Fund for China Agriculture Research System under Grant CARS-45-39; Hangzhou Agricultural & Social Development Research Program under Grant 20162012A03; Science & Technology Innovation Program of Hangzhou Academy of Agricultural Sciences under Grant 2019HNCT-01.

Author Contribution

Kai Liu and Nan Xie conducted the experiments; Kai Liu analyzed the data and wrote the manuscript; Xiao-yu Feng and Heng-jia Ma contributed to data analysis and revising the manuscript; Nan Xie supervised the whole project.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper. The authors alone are responsible for the content and writing of the paper.

Acknowledgments

This work was supported by the Earmarked Fund for China Agriculture Research System under Grant CARS-45-39; Hangzhou Agricultural & Social Development Research Program under Grant 20162012A03; Science & Technology Innovation Program of Hangzhou Academy of Agricultural Sciences under Grant 2019HNCT-01. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors alone are responsible for the content and writing of the paper.

References

Balakrishnan, R., Harris, M. A., Huntley, R., Van Auken, K., & Cherry, J. M. (2013). A guide to best practices for Gene Ontology (GO) manual annotation. *Database: the journal of biological databases and curation*, 2013, bat054. https://doi.org/10.1093/database/bat054

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32, 314-331.

Bouza, C., Hermida, M., Pardo, B. G., Vera, M., Fernández, C., de la Herrán, R., Navajas-Pérez, R., Álvarez-Dios, J. A., Gómez-Tato, A., & Martínez, P. (2012). An Expressed Sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): a useful framework for comparative genomics across model and farmed teleosts. *BMC genetics*, 13, 54. https://doi.org/10.1186/1471-2156-13-54

Chistiakov, D. A., Hellemans, B. & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*, 255, 1-29. https://doi.org/10.1016/j.aquaculture.2005.11.031

Du, R., Zhang, D. L., Wang, Y. Z., Wang, W. M., & Gao, Z. X. (2016). Cross-species amplification of microsatellites in genera *Megalobrama* and *Parabramis*. *Journal of Genetics*, 93(Suppl 2), 106-109. https://doi.org/10.1007/s12041-013-0308-1

Dakin, E. E., & Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. *Heredity* 93, 504-509. https://doi.org/10.1038/sj.hdy.6800545

De Barba, M., Miquel, C., Lobreaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, 17(3), 492-507. https://doi.org/10.1111/1755-0998.12594

Ellis, J. R., & Burke, J. M. (2007). EST-SSRs as a resource for population genetic analyses. Heredity, 99(2), 125–132. https://doi.org/10.1038/sj.hdy.6801001

Froese, R., & Pauly, D. (2019). *FishBase*. https://www.fishbase.se/summary/Megalobrama-terminalis.html.

Feng, X., Yu, X., Fu, B., Wang, X., Liu, H., Pang, M., & Tong, J. (2018). A high-resolution genetic linkage map and QTL fine mapping for growth-related traits and sex in the

Yangtze River common carp (*Cyprinus carpio haematopterus*). *BMC genomics*, 19(1), 230. https://doi.org/10.1186/s12864-018-4613-1

Gao, Z., Luo, W., Liu, H., Zeng, C., Liu, X., Yi, S., & Wang, W. (2012). Transcriptome analysis and SSR/SNP markers information of the blunt snout bream (*Megalobrama amblycephala*). *PLos One*, 7(8), e42637. https://doi.org/10.1371/journal.pone.0042637

Ghebranious, N., Vaske, D., Yu, A., Zhao, C. F., Marth, G., & Weber, J. L. (2003). STRP screening sets for the human genome at 5 cM density. *BMC Genomics*, 4(1), 6. https://doi.org/10.1186/1471-2164-4-6

Green, M. R., & Sambrook, J. (2012). *Molecular Cloning*: *A Laboratory Manual*. *4th Edition*. New York: Cold Spring Harbor Laboratory.

Hasselman, D. J., Ricard, D., & Bentzen, P. (2013). Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular Ecology*, 22, 1558-1573. https://doi.org/10.1111/mec.12197

Hill, D.P., Smith, B., McAndrews-Hill, M.S. & Blake, J. A. Gene Ontology annotations: what they mean and where they come from. *BMC Bioinformatics* 9, S2 (2008). https://doi.org/10.1186/1471-2105-9-S5-S2

Hu, X. S., & Shi, L. Y. (2020). A Review: Research Progress on Germplasm Resource of Black Bream (*Megalobrama terminalis*) in China. *Chinese Journal of Fisheries*, 33(3): 84-89.

Johnson, M. S., & Black, R. (1984). The Wahlund effect and geographical scale of variation in the intertidal limpet *Siphonaria* sp. *Marine Biology*, 79, 295-302. https://doi.org/10.1007/BF00393261

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increase success in paternity assignment. *Molecular Ecology*, 16, 1099-1106. https://doi.org/10.1111/j.1365-294x.2007.03089.x

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, 10(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Liu, B., Kuang, Y. Y., Tong, G. X., & Yin, J. S. (2011). Analysis of genetic diversity on 9 wild stocks of Taimen (*Hucho taimen*) by microsatellite markers. *Zoological Research*, 32, 597-604. https://doi.org/10.3724/sp.j.1141.2011.06597

Liu, H. G., Yang, Z., Tang, H. Y., Gong, Y., & Wan, L. (2017). Microsatellite development and characterization for *Saurogobio dabryi* Bleeker, 1871 in a Yangtze river-connected lake, China. *Journal of Genetics*, 96, e1-e4. https://doi.org/10.1007/s12041-016-0733-z

Liu, K., Xie, N., Feng, X. Y., & Ma, H. J. (2020). Characterization of 15 polymorphic microsatellite markers developed by magnetic beads enrichment in *Megalobrama terminalis*. *Journal of Zhejiang Agricultural Sciences*, 61(1), 144-146. https://doi.org/10.16178/j.issn.0528-9017.20200144

Lopez, L., Barreiro, R., Fischer, M., & Koch, M. A. (2015). Mining microsatellite markers from public expressed sequence tags databases for the study of threatened plants. *BMC Genomics*, 16, 781. https://doi.org/10.1186/s12864-015-2031-1

Ma, K. Y., Feng, J. B., Xie, N., Feng, X. Y., & Li, J. L. (2011). Microsatellite analysis of genetic variation of the oriental river prawn *Macrobrachium nipponense* in Qiantang River. *Zoological Research*, 3, 363-370. https://doi.org/10.3724/sp.j.1141.2011.04363

Manco, R., Chiaiese, P., Basile, B., & Corrado, G. (2020). Comparative analysis of genomic- and EST-SSRs in European plum (Prunus domestica L.): implications for the diversity analysis of polyploids. 3 Biotech, 10(12), 543. https://doi.org/10.1007/s13205-020-02513-w

Parthiban, S., Govindaraj, P., & Senthilkumar, S. (2018). Comparison of relative efficiency of genomic SSR and EST-SSR markers in estimating genetic diversity in sugarcane. 3 Biotech, 8(3), 144. https://doi.org/10.1007/s13205-018-1172-8

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, 43, 223-225. https://doi.org/10.1111/j.1558-5646.1989.tb04220.x

Shackel, N. A., Gorrell, M. D., & McCaughan, G. W. (2002). Gene array analysis and the liver. *Hepatology*, 36(6), 1313-1325. https://doi.org/10.1053/jhep.2002.36950

Shackel, N. A., Seth, D., Haber, P. S., Gorrell, M. D., & McCaughan, G. W. (2006). The hepatic transcriptome in human liver disease. *Comparative Hepatology*, 5, 6. https://doi.org/10.1186/1476-5926-5-6

Slatkin, M., & Excoffier, L. (1996). Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity*, 76 (4), 377–383. https://doi.org/10.1038/hdy.1996.55

Song, W., Zhu, D., Lv, Y., & Wang, W. M. (2016). Development and characterization of 29 polymorphic microsatellite loci of *Megalobrama pellegrini* by next-generation sequencing technology and cross-species amplification in related species. *PeerJ Preprints*, 4, e2490v1. https://doi.org/10.7287/peerj.preprints.2490v1

Song, W., Zhu, D., Lv, Y., & 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. *Journal of Genetics*, 96, 39-45. https://doi.org/10.1007/s12041-017-0815-6

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources*, 4, 535-538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Vieira, M. L., Santini, L., Diniz, A. L., & Munhoz, C. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology*, 39(3), 312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Weber, J. L., & Broman, K. W. (2001). Genotyping for human whole-genome scans: past, present, and future. *Advances in Genetics*, 42, 77-96. https://doi.org/10.1016/s0065-2660(01)42016-5

Xu, J., Liu, L., Xu, Y., Chen, C., Rong, T., Ali, F., Zhou, S., Wu, F., Liu, Y., Wang, J., Cao, M., & Lu, Y. (2013). Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA research*, 20(5), 497–509. https://doi.org/10.1093/dnares/dst026

Yeh, F. C., Yang, R. C., Boyle, T. B. J., Ye, Z. H., & Mao, J. X. (1999). *POPGENE 1.32*, *the user-friendly shareware for population genetic analysis*. Edmonton: Molecular Biology and Biotechnology Center, University of Alberta. https://sites.ualberta.ca/~fyeh/popgene_download.html.

Yoshida, G. M., Barria, A., Correa, K., Cáceres, G., Jedlicki, A., Cadiz, M. I., Lhorente, J. P., & Yáñez, J. M. (2019). Genome-Wide Patterns of Population Structure and Linkage Disequilibrium in Farmed Nile Tilapia (*Oreochromis niloticus*). *Frontiers in genetics*, 10, 745. https://doi.org/10.3389/fgene.2019.00745

Yu, J. N., Kim, S. K., Sagong, J., Ryu, S. H., & Chae, B. (2019). Identification of microsatellite markers and their application in yellow catfish (*Pseudobagrus fulvidraco* Richardson, 1846) population genetics of Korea. *Journal of Genetics*, 98, 2-7. https://doi.org/10.1007/s12041-018-1047-0

Zhang, L., Yuan, D., Yu, S., Li, Z., Cao, Y., Miao, Z., Qian, H., & Tang, K. (2004). Preference of simple sequence repeats in coding and non-coding regions of Arabidopsis thaliana. *Bioinformatics (Oxford, England)*, 20(7), 1081–1086. https://doi.org/10.1093/bioinformatics/bth043

Zhang, Q. Q., Chen, J., Jiang, X. Y., & Zou, S. M. (2014). Establishment of DNA fingerprinting and analysis on genetic structure of different *Parabramis* and *Megalobrama* populations with microsatellite. *Journal of Fisheries of China*, 38, 15-23.

Zheng, X. F., Pan, C., Diao, Y., You, Y. N., Yang, C. Z., & Hu, Z. L. (2013). Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Aracease). *BMC Genomics*, 14, 490. https://doi.org/10.1186/1471-2164-14-490

Zhao, Y., Ji, X. S., Zeng, Y. Q., Ding, L., Yang, P. P., & Wang, H. (2011). Isolation of microsatellite markers for *Lateolabrax japonicus* and polymorphic analysis. *Zoological Research*, 32, 515-520. https://doi.org/10.3724/sp.j.1141.2011.05515