Development and characterization of 29 polymorphic microsatellite loci of *Megalobrama pellegrini* by next-generation sequencing technology and cross-species amplification in related species

Wen Song [1] , **Dongmei Zhu** [1] , **Yefeng Lv** [1] , **Weimin Wang** Corresp. [1]

[1] Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, Huazhong Agricutural University, Wuhan, China

Corresponding Author: Weimin Wang
Email address: wangwm@mail.hzau.edu.cn

Megalobrama pellegrini is one of the economically important freshwater fish in China. Here, we developed 29 polymorphic microsatellite loci of *M. pellegrini*. The number of alleles (*Na*), effective number of alleles (*Ne*), observed heterozygosity (H_O), expected heterozygosity (H_E) and polymorphic information content (PIC) ranged from 3 to 11 (mean±SD 5.4828±1.9571), 2.8708 to 9.6257 (mean±SD 5.0865±1.6681), 0.4333 to 0.9333 (mean±SD 0.7874±0.1213), 0.6627 to 0.9113 (mean±SD 0.7946±0.0751) and 0.5785 to 0.8868 (mean±SD 0.7439±0.0950), respectively. Cross-species amplification was successful at most loci for related species such as *M. amblycephala*, *M. hoffmanni*, *M. skolkovii* and *Parabramis pekinensis*. The transferability rate of the 29 polymorphic microsatellite markers in *M. amblycephala*, *M. hoffmanni*, *M. skolkovii* and *P. pekinensis* were 96.55%, 86.21%, 86.21% and 75.86%, respectively. These polymorphic microsatellites are not only useful in genetic study and conservation of *M. pellegrini*, but also an effective tool for identifying the related species. We could use 5 microsatellite markers (HHF-63, HHF-104, HHF-113, HHF-148, HHF-163) to distinguish the 5 species.

Development and characterization of 29 polymorphic microsatellite loci of *Megalobrama pellegrini* by next-generation sequencing technology and cross-species amplification in related species

Wen Song, Dongmei Zhu,Yefeng Lv and Weimin Wang

Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, College of Fisheries, Huazhong Agricultural University, Wuhan 430070, China

Corresponding Author: Weimin Wang

wangwm@mail.hzau.edu.cn

Abstract

Megalobrama pellegrini is one of the economically important freshwater fish in China. Here, we developed 29 polymorphic microsatellite loci of *M. pellegrini*. The number of alleles (*Na*), effective number of alleles (*Ne*), observed heterozygosity (H_O), expected heterozygosity (H_E) and polymorphic information content (PIC) ranged from 3 to 11 (mean ±SD 5.4828 ±1.9571), 2.8708 to 9.6257 (mean ±SD 5.0865 ±1.6681), 0.4333 to 0.9333 (mean ±SD 0.7874 ±0.1213), 0.6627 to 0.9113 (mean ±SD 0.7946 ±0.0751) and 0.5785 to 0.8868 (mean ±SD 0.7439 ±0.0950), respectively. Cross-species amplification was successful at most loci for related species such as *M. amblycephala*, *M. hoffmanni*, *M. skolkovii* and *Parabramis pekinensis*. The transferability rate of the 29 polymorphic microsatellite markers in *M. amblycephala, M. hoffmanni, M. skolkovii and P. pekinensis* were 96.55%, 86.21%, 86.21% and 75.86%, respectively. These polymorphic microsatellites are not only useful in genetic study and conservation of *M. pellegrini*, but also an effective tool for identifying the related species. We could use 5 microsatellite markers (HHF-63, HHF-104, HHF-113, HHF-148, HHF-163) to distinguish the 5 species.

Key words: *Megalobrama pellegrini*, *Megalobrama hoffmanni*, *Megalobrama skolkovii*, *Megalobrama amblycephala*, *Parabramis pekinensis*, Microsatellite, Polymorphism

Introduction

Microsatellites, also known as simple sequence repeats, have been extensively used as molecular markers due to their abundance and high degree of polymorphism (Ellegren, 2004; Saenz-Agudelo et al., 2015; Camargo et al., 2015). With the development of next-generation sequencing technology, the sequencing throughput has dramatically increased and the costs have decreased. Microsatellites can be developed quickly and at low cost for most species using next-generation sequencing in recent years (Gardner et al., 2011; Berman, Austin & Miller, 2014; Plough & Marko, 2014). Microsatellites are currently the most widely used molecular markers in genetic breeding, genome mapping, gene location, species identifying and so on.

Megalobrama pellegrini belongs to the genus *Megalobrama* that also includes *M. amblycephala*, *M. hoffmanni* and *M. skolkovii*, and is one of the most important economic

freshwater fish in China (Wu, 1964; Luo, 1990; Chen, 1998). The species of the genus *Megalobrama* are extremely similar in morphological traits (Song et al., 2013). *Parabramis pekinensis* is similar in morphological traits with the species of the genus *Megalobrama* (Cai et al., 2001). *M. pellegrini* distributes in the upstream area of Yangtze River that differs from the other species of the genus *Megalobrama* (Chen, 1998). Due to habitat changes and human factors such as pollution and overfishing, populations of *M. pellegrini* have been decreasing rapidly across their original distribution (Chen, 2002; Xie, 2003; Gao et al., 2010). Since the 21st century, the natural resources and population of *M. pellegrini* have obviously declined (Li et al., 2005; Gao et al., 2009). *M. pellegrini* is hard to find in large tributaries of Yangtze River, and they only distribute in some small flow tributaries. *M. pellegrini* has become an endangered species, so that it is in need of protection (Wang et al., 2012). But there are only a few articles about *M. pellegrini*, and only one article is about microsatellite markers of *M. pellegrini* (Wang et al., 2012). Here, we developed 29 polymorphic microsatellite loci for *M. pellegrini* by next-generation sequencing technology. These polymorphic microsatellite markers will be useful in protecting genetic resources and molecular breeding of this species. They also offers the potential to identify species of *M. pellegrini*, *M. hoffmanni*, *M. skolkovii*, *M. amblycephala* and *P. pekinensis*.

Materials & methods

Aquaculture materials and genomic DNA extraction

The sampling positions of the 5 species were listed in Table 1. 16 individuals of the population of *M. pellegrini* were used for the initial screening of microsatellite loci. The DNA of the *M. pellegrini* was extracted using the CTAB method (Cota-Sánchez, Remarchuk & Ubayasena, 2006). The concentration of each DNA sample was measured using a spectrophotometer at 260 and 280 nm and each DNA sample was quantified on an agarose gel. Genome sequencing of *M. pellegrini* was performed on an Illumina HiSeq 2000™ by Biomarker Technologies (Beijing, China). The procedure was performed in accordance with the standard Illumina protocol, including sample preparation and sequencing as follows: quantified DNA was extracted and DNA fragments were obtained using ultrasound, which was purified using the QIA quick PCR kit. End repair was performed with poly-A on the 3'

ends, then the adaptors were ligated, clusters were generated, agarose gel electrophoresis was used to select fragments, and PCR amplification was performed. Sequencing was performed by establishing a library with Illumina HiSeq 2000^{TM}. The short reads were aligned using the Burrows-Wheeeler transformation (Li & Durbin, 2010). The microsatellite primer pairs of *M. pellegrini* were designed using Primer Premier 6.0 software, which was used to check against potential primer dimers, hairpin structures and the occurrence of mismatches. The specificity of primer pairs were checked by comparing their sequences with the genomic sequences. Only those microsatellite loci that contained motifs 4 to 6 nucleotides in size were selected to synthesize. The minimum motif repeat was defined as 5 for hexanucleotides, 10 for pentanucleotides and 20 for tetranucleotides. Parameters for designing the primers were set as follows: primer length ranged from 20 bases to 28 bases with 24 as the optimum; PCR product size ranged from 160-300 bp; melting temperature ranged from 55℃ to 63℃ with 59℃ as the optimum annealing temperature; GC content ranged from 40% to 70% with 50% as the optimum. At last, total of 192 primer pairs were chosen to synthesize.

PCR amplification and microsatellite loci analysis

These primer pairs, which were successfully amplified and found to be polymorphic in initial screening tests, were then tested in 30 individuals. The PCR amplification reactions were performed using a thermal cycler (Eppendof, Germany). The PCR was performed in a final volume of 20 μL. The PCR mixture was as follows: 2 μl genomic DNA (100 ng/μl), 10 μl PCR mix (TaKaRa, Japan), 1 μl of each primer (10 μM) and 6 μl ddH_2O. The PCR conditions was as follows: 95℃ for 5 min, 40 cycles with 95℃ denaturation for 30s, 72℃ annealing for 45s, and 72℃ elongation for 10 min，followed by 4℃ for the end. Each effective PCR product was validated in 30 individuals of *M. pellegrini* and analyzed using QIAxcel Advanced (Qiagen, Germany). Population genetic parameters for polymorphic loci such as number of alleles (*Na*), effective number of alleles (*Ne*), observed heterozygosity (H_O), expected heterozygosity (H_E), polymorphic information content (PIC), *P* values in Hardy-Weinberg equilibrium (HWE) tests were calculated by GENEPOP 4.0 (Rousset, 2007). Finally, the polymorphic primer pairs were assessed for cross-amplification in other 4 related species (*M. amblycephala*, *M. hoffmanni*, *M. skolkovii*, *P. pekinensis*) each in 30 individuals,

using the PCR reactions as described above.

Results

Frequency and distribution of microsatellite loci in the genome

In this study, total of 164571 sequences that contained microsatellite loci could be used to develop microsatellite markers from genomic DNA of *M. pellegrini*. The length of the microsatellite loci ranged from 12 to 248 bp (mean ±SD 19.49 ±13.54). The microsatellite loci included 5 motif types: dinucleotide repeats (48.15%), trinucleotide repeats (31.74%), tetranucleotide repeats (16.49%), pentanucleotide repeats (3.31%) and hexanucleotide repeats (0.31%) (Fig. 1 A). The motif repeat number of microsatellite loci ranged from 4 to 114, while 98.74% of the microsatellite loci had 4-30 motif repeats, and motifs with more than 50 repeats were only with a frequency of <0.5% (Fig. 1 B). There were 8 types of dinucleotide motifs: CA/TG (29.08%), AC/GT (28.87%), AG/CT (12.61%), GA/TC (12.28%), AT (8.73%), TA (8.25%), GC (0.11%) and CG (0.08%) (Fig. 1 C). The number of trinucleotide motif types was 30, and the main types respectively were: AAT/ATT (20.16%), TAA/TTA (17.65%) and ATA/TAT (10.85%) (Fig. 1 D). The number of tetranucleotide motif types was 116, and the main types respectively were: AAAT/ATTT (16.79%), TAAA/TTTA (11.05%) and AATA/TATT (6.47%) (Fig. 1 E). The number of pentanucleotide motif types was 371, and the main types respectively were: AAAAT/ATTTT (5.53%), TAAAA/TTTTA (5.10%) and CAAAA/TTTTG (4.42%) (Fig. 1 F). The number of hexanucleotide motif types was 295, and the main types respectively were: TAAATG/CATTTA (3.48%), AAAAAT/ATTTTT (3.29%) and AAAAAG/CTTTTT (2.32%) (Fig. 1 G).

Polymorphism and transferability of microsatellite markers

Here, 192 of 164571 primer pairs were chosen to synthesize. The successful amplification rate of these primer pairs was high, 124 pairs successfully amplified genomic DNA of *M. pellegrini*. Among the 124 pairs, 97 pairs generated target bands. The rest pairs were not suitable that they either generated non-target bands or the produced amplified bands were too weak. 32 of the 97 primer pairs were found to be polymorphic, with 30 individuals being used as PCR templates. Among the polymorphic primer pairs, 3 pairs were excluded for

further analysis because their amplified fragments exhibited an excess of stutter bands preventing the unambiguous identification of alleles. Finally, we got 29 polymorphic primer pairs (Table 2).The *Na*, *Ne*, H_O, H_E and PIC ranged from 3 to 11 (mean±SD 5.4828±1.9571), 2.8708 to 9.6257 (mean±SD 5.0865±1.6681), 0.4333 to 0.9333 (mean±SD 0.7874±0.1213), 0.6627 to 0.9113 (mean±SD 0.7946±0.0751) and 0.5785 to 0.8868 (mean±SD 0.7439±0.0950), respectively. Of the 29 loci, 3 loci (HHF-63，HHF-148，HHF-157) deviated significantly from HWE (P<0.05) (Table 2).

The 29 polymorphic microsatellite markers were used to amplify in 4 species: *M. amblycephala, M. hoffmanni, M. skolkovii and P. pekinensis*, and the transferability rates were 96.55%, 86.21%, 86.21% and 75.86%, respectively (Table 3). In *M. amblycephala*, 28 loci were amplified successfully, only 1 loci (HHF-104) produced no band after amplification. Among the 28 loci, 22 loci were monomorphic, and 6 loci (HHF-66, HHF-84, HHF-103, HHF-108, HHF-163 and HHF-177) were polymorphic in *M. amblycephala*. The *Na*, *Ne*, H_O, H_E, PIC ranged from 2 to 4 (mean±SD 2.8333±0.7528), 1.9231 to 3.9301 (mean±SD 2.7891±0.7528), 0.4667 to 0.8333 (mean±SD 0.6278±0.1254), 0.4881 to 0.7582 (mean±SD 0.6283±0.1092) and 0.3648 to 0.6981 (mean±SD 0.5336±0.1353), respectively. The result of the other 3 species was showed in Table 3.

Discussion

Due to low costs and reduced expenditure of time, next-generation sequencing technology become a powerful tool to develop microsatellite markers (Taguchi et al., 2013; Thai et al., 2016; Wang et al., 2016). Next-generation sequencing technology is presented to construct a genomic DNA sequence library highly enriched for microsatellite sequences. Total of 164571 microsatellite sequences were used to design primers, and 192 primer pairs were chosen to synthesize. There were 29 polymorphic microsatellite loci isolated and amplified, and 3 loci among them deviated significantly from HWE (P< 0.05), which indicated there might be presence of null alleles in these loci (Zeng et al., 2013). The polymorphic loci were assessed to contain moderately high polymorphism degree (PIC > 0.50) (Botstein et al., 1980), so that they would be potentially useful for applications in population genetics and molecular ecology of *M. pellegrini*. The microsatellite markers were genetic analysis tools to manage

breeding programs of cultured species. In contrast, Wang et al (2012) developed 26 polymorphic microsatellite markers (the mean *Na*, H_O and H_E were 3.08, 0.47 and 0.51), the microsatellite loci we developed showed higher polymorphism.

Furthermore, cross-species amplification was tested using the 29 polymorphic primer pairs within the 4 species: *M. amblycephala, M. hoffmanni, M. skolkovii* and *P. pekinensis*. The amplification rate with related species was high, and corresponded with the phylogenetic distance between species, varying from 75.86% in *P. pekinensis* to 96.55% in *M. amblycephala*. The 5 species were very similar in morphological traits, especially the species of the genus *Megalobrama* (Li et al., 2001; Chen et al., 2014). We could distinguish the 5 species (*M. pellegrini*, *M. amblycephala, M. hoffmanni, M. skolkovii* and *P. pekinensis*) according to characteristics of 29 microsatellite markers in cross-species amplifications (Table 3). We could use 5 microsatellite markers (HHF-63, HHF-104, HHF-113, HHF-148, HHF-163) to identify the 5 species: when amplifications using HHF-104 and HHF-148 produced band, we recognize it as *M. pellegrini*; when amplification using HHF-104 produce no band and amplification using HHF-113 produce band, we recognize it as *M. amblycephala*; when both amplifications using HHF-63 and HHF-163 produce band and amplifications using HHF-113 produce no band, we recognize it as *M. skolkovii*; when amplifications using HHF-163 produced no band, we recognize it as *M. hoffmanni*; when amplifications using HHF-63 produced no band, we recognize it as *P. pekinensis*. Some other marker combinations of the 29 polymorphic microsatellite loci also can be used to identify the 5 species (Table 2, Table 3). Cross-species amplification was very significant that it could improve the utilization of microsatellite markers. Qu et al (2012) used 46 polymorphic microsatellite markers of two sinipercine fishes (*Siniperca*) to amplify in other 4 species of sinipercine fishes, and the interspecies cross-amplification rate was 94%. Wang et al (2015) used polymorphic microsatellite markers of *Hucho bleekeri* to amplify in related species. In fact, cross-species amplification of microsatellite markers could be used in related species identification. The classification of the genus *Megalobrama* has always been a controversy (Luo, 1990), and results of cross-species amplification would contribute to classification of the genus *Megalobrama*.

In conclusion, we developed 29 polymorphic microsatellite markers for *M. pellegrini*. These

markers will be useful for conservation of *M. pellegrini*. Moreover, these polymorphic microsatellite markers are helpful for identifying related species and have potential usefulness for comparative genetic studies across related species.

Acknowledgements

This work was supported by the Modern Agriculture Industry Technology System Construction Projects of China titled as "Staple Freshwater Fishes Industry Technology System" (No. CARS-46-05).

References

Berman M, Austin CM, Miller AD. 2014. Characterisation of the complete mitochondrial genome and 13 microsatellite loci through next-generation sequencing for the New Caledonian spider-ant *Leptomyrmex pallens*. *Molecular Biology Reports* 41: 1179-1187. DOI: 10.1007/s11033-013-2657-5.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics* 32: 314-331.

Cai MJ, Zhang MY, Zeng QL, Liu HZ. 2001. A study on the morphological of the genus *Megalobrama*. *Acta Hydrobiologica Sinica* 25: 631-635.

Camargo C, Costa MC, Del Rio GC, Gibbs HL, Glenn TC, Bagal U, Silveira LF, Wasko AP, Francisco MR. 2015. Novel and cross-amplified microsatellite loci for the critically endangered São Paulo marsh antwren *Formicivora paludicola* (Aves: Thamnophilidae). *Conservation Genetic Resources* 7: 129-131. DOI: 10.1007/s12686-014-0310-9.

Chen C. 2002. The impact of dams on fisheries: Case of the Three Gorges Dam *challenges of a changing earth*. Berlin: Springer Verlag, 97-99.

Chen J, Li FG, Huang CX, Jiang XY, Zou SM. 2014. Morphological variations of genera *Parabramis* and *Megalobrama* teleost populations. *Journal of Shanghai University* 23: 388-394.

Chen YY. 1998. *Fauna Sinica, Osteichthyes Cypriniformes II*. Beijing: Science Press, 198-207.

Cota-Sánchez JH, Remarchuk K, Ubayasena K. 2006. Ready-to-use DNA extracted with a CTAB method adapted for herbarium specimens and mucilaginous plant tissue. *Plant Molecular Biology Reporter* 24: 161-167. DOI: 10.1007/bf02914055.

Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435-445. DOI:10.1038/nrg1348.

Gao X, Tan DQ, Liu HZ, Wang JW. 2009. Exploitation status and conservation of a population of *Megalobrama pellegrini* in Longxi river in the upper Yangtze River basin. *Sichuan Journal of Zoology* 28: 329-333.

Gao X, Zeng Y, Wang J, Liu HZ. 2010. Immediate impacts of the second impoundment on fish communities in the Three Gorges Reservoir. *Environmental Biology of Fishes* 87: 163-173. DOI:10.1007/s10641-009-9577-1.

Gardner MG, Fitch AJ, Bertozzi T, Lweo AJ. 2011. Rise of the machines–recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* 11: 1093-1101. DOI: 10.1111/j.1755-0998.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754-1760. DOI: 10.1093/bioinformatics/btp324.

Li SF, Zhu ZW, Zou SM, Zhao JL, Cai WQ. 2001. Interspecific phylogenesis and intraspecific genetic differences of genus *Megalobrama*: bluntnose black bream (*M. amblycephala*), guangdong black bream (*M. hoffmanni*) and black bream (*M. terminalis*). *Acta zoologica Sinica* 48: 339-345.

Li WJ, Wang JW, Tan DQ, Dan SG. 2005. Observation on postembryonic development of *Megalobrama pellegrini*. *Journal of Fisheries of China* 6: 729-736.

Luo YL. 1990. A revision of fishes of the cyprinid genus *Megalobrama*. *Acta Hydrobiologica Sinica* 14: 160-165.

Plough LV, Marko PB. 2014. Characterization of microsatellite loci and repeat density in the gooseneck barnacle, *Pollicipes elegans*, using next generation sequencing. *Journal of Heredity* 105: 136-142. DOI: 10.1093/jhered/est064.

Qu CM, Liang XF, Huang W, Cao L. 2012. Isolation and characterization of 46 novel polymorphic EST-simple sequence repeats (SSR) markers in two Sinipercine fishes (*Siniperca*) and cross-species amplification. *International Journal of Molecular Sciences* 13:

9534-9544. DOI:10.3390/ijms13089534.

Rousset F. 2007. GENEPOP'007: A complete reimplementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103-106. DOI: 10.1111/j.1471-8286.

Saenz-Agudelo P, Almany GR, Mansour H, Perumal S, Berumen ML. 2015. Characterization of 11 novel microsatellite markers for the vagabond butterfly fish, *Chaetodon vagabundus*. *Conservation Genetic Resources* 7: 713-714. DOI: 10.1007/s12686-015-0440-8.

Song W, Wang YZ, Zhu DM, Ren L. 2013. Morphological variations among the genus *Megalobrama*. *Freshwater Fisheries* 43: 21-27.

Taguchi M, Shigenobu Y, Ohkubo M, Yanagimoto T, Sugaya T, Nakamura Y, Saitoh K, Yokawa K. 2013. Characterization of 12 polymorphic microsatellite DNA loci in the blue shark, *Prionace glauca*, isolated by next generation sequencing approach. *Conservation Genetic Resources* 5: 117-119. DOI: 10.1007/s12686-012-9746-y.

Thai BT, Tan MH, Lee YP, Gan HM, Tran TT, Austin C. 2016. Characterisation of 12 microsatellite loci in the Vietnamese commercial clam *Lutraria rhynchaena* Jonas 1844 (Heterodonta: Bivalvia: Mactridae) through next-generation sequencing. *Molecular Biology Reports* 43: 391-396. DOI: 10.1007/s11033-016-3966-2.

Wang JJ, Yu XM, Zhao K, Zhang YG, Tong JG, Peng ZG. 2012. Microsatellite development for an endangered bream *Megalobrama pellegrini* (Teleostei, Cyprinidae) using 454 sequencing. *International Journal of Molecular Sciences* 13: 3009-3021. DOI: 10.3390/ijms13033009.

Wang K, Zhang SH, Wang DQ, Xin MM, Wu JM, Sun QL, Du H, Wang CY, Huang J, Wei QW. 2015. Development of 27 novel cross-species microsatellite markers for the endangered *Hucho bleekeri* using next-generation sequencing technology. *Conservation Genetic Resources* 7: 263-267. DOI: 10.1007/s12686-014-0353-y.

Wang YZ, Cao LJ, Zhu JY, Wei SJ. 2016. Development and characterization of novel microsatellite markers for the peach fruit moth *Carposina sasakii* (Lepidoptera: Carposinidae) using next-generation sequencing. *International Journal of Molecular Sciences* 17: 362. DOI: 10.3390/ijms17030362.

Wu XW. 1964.*Ichthyography of Cyprinidae in China* (DownVolume). Shanghai: Shanghai

Science and Technology Press, 93-118.

Xie P. 2003. Three-Gorges Dam: risk to ancient fish. *Science* 302: 1149-1151. DOI: 10.1126/science.302.5648.1149b.

Zeng Q, Ye H, Ludwig A, Wang Z, Zhang Y, Peng Z. 2013. Microsatellite development for the endangered Yangtze sturgeon (*Acipenser dabryanus* Duméril, 1869) using 454 sequencing. *Journal of Applied Ichthyology* 29: 427-514. DOI: 10.1111/jai.12278.

Table 1(on next page)

Sampling details of *M. pellegrini*, *M. amblycephala*, *M. hoffmanni*, *M. skolkovii* and *P. pekinensis*

Table 1 Sampling details of *M. pellegrini*, *M. amblycephala*, *M. hoffmanni*, *M. skolkovii* and *P. pekinensis*

Species	Location	Latitude (N)	Longitude(E)	Altitude (m)	Sample size
M. pellegrini	Longxi Lake, Luzhou, Sichuan Province	29°03'	105°51'	318	30
M. amblycephala	Liangzi Lake, Ezhou, Hubei Province	30°11'	114°64'	20	30
M. hoffmanni	Xijiang River, Zhaoqing, Guangdong Province	23°04'	112°47'	0	30
M. skolkovii	Heilong River, Fuyuan, Heilongjiang Province	48°37'	134°28'	33	30
P. pekinensis	Yangtze River, Huanggang, Hubei Province	30°44'	114°86'	21	30

Figure 1(on next page)

Characterization of microsatellite loci in *M. pellegrini* genome

A Distribution of different SSR repeat motif types; B Number of different repeat motifs; C Frequency distribution of dinucleotide repeats based on different motif types; D Frequency distribution of trinucleotide repeats based on different motif types; E Frequency distribution of tetranucleotide repeats based on main motif types (>1.00%); F Frequency distribution of pentanucleotide repeats based on main motif types (>1.00%); G Frequency distribution of hexanucleotide repeats based on main motif types (>0.70%)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2490v1 | CC BY 4.0 Open Access | rec: 30 Sep 2016, publ: 30 Sep 2016

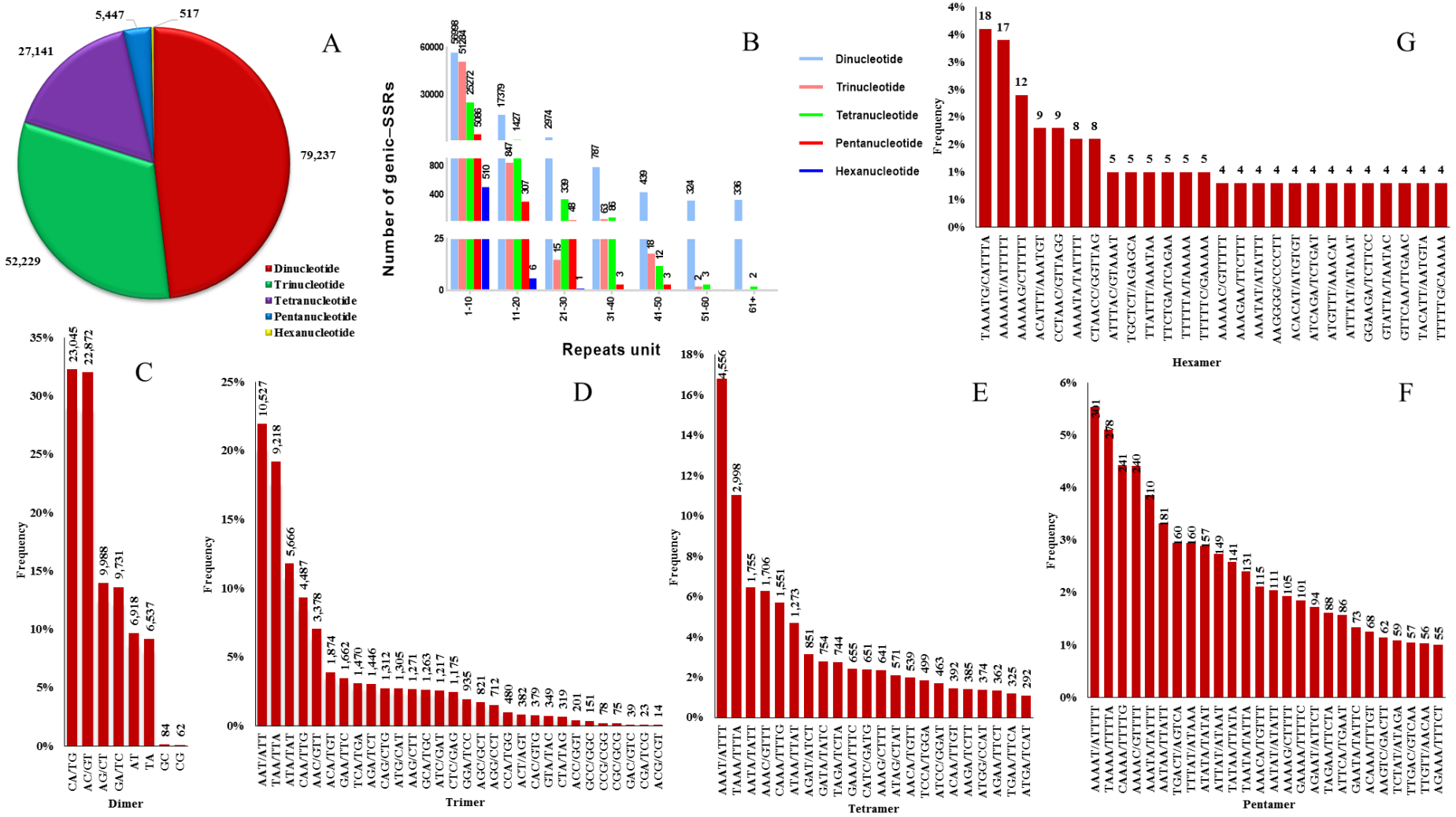

Figure 1 Characterization of microsatellite loci in *M. pellegrini* genome. A Distribution of different SSR repeat motif types; B Number of

different repeat motifs; C Frequency distribution of dinucleotide repeats based on different motif types; D Frequency distribution of trinucleotide repeats based on different motif types; E Frequency distribution of tetranucleotide repeats based on main motif types (>1.00%); F Frequency distribution of pentanucleotide repeats based on main motif types (>1.00%); G Frequency distribution of hexanucleotide repeats based on main motif types (>0.70%)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2490v1 | CC BY 4.0 Open Access | rec: 30 Sep 2016, publ: 30 Sep 2016

Table 2(on next page)

Characteristics of 29 microsatellite loci developed for *M. pellegrini*

Table 2 Characteristics of 29 microsatellite loci developed for *M. pellegrini*

Locus	Genbank Accession No	Primer sequence (5′to 3′)	Repeat motif	Size range (bp)	Tm℃	*Na*	*Ne*	H_O	H_E	PIC	HWE *P* value
HHF-5	KX548362	F: CAAACTATTGGAGGGCAGTGTATTTT R: GTTCATCTTGGCCCCTATTTTTGC	$(AATAAG)_8$	217-277	56.0	5	4.4776	0.9000	0.7898	0.7413	0.4183
HHF-22	KX548363	F: CGGATTTTCACCTGCTCTGCCTTT R: CCTGAACTCTTTCCTGCTCAGATCTG	$(CACACG)_5$	202-214	60.0	3	2.8708	0.6667	0.6627	0.5785	0.8501
HHF-42	KX548364	F: CTGTATTGCACAAACTCCCATGAAGT R: CAGAGTCACCTGGAATCACTCTGC	$(TTTTG)_{12}$	259-284	57.5	6	5.8065	0.9000	0.8418	0.8038	0.3256
HHF-58	KX548365	F: CACAGAGGAGTTCACATGAGTGTT R: TCTTTTTCAGGCACCAAGGAAATCT	$(TAGAA)_{14}$	215-245	55.8	7	6.6176	0.7333	0.8633	0.8300	0.2034
HHF-61	KX548366	F: TTTTGGGTAATGCAGCCCTGTTCA R: CCGAAATCCCACCTCAATAGCTGA	$(AGAAT)_{13}$	270-295	56.0	6	5.5728	0.8667	0.8345	0.7951	0.2135
HHF-62	KX548367	F: GGCTTTGGACCAGACTTATAAGAGCT R: CGCCCACACTAACCCATGATGTAT	$(CATAA)_{12}$	210-240	55.0	7	6.6176	0.9333	0.8633	0.8300	0.7367
HHF-63	KX548368	F: ACTGACACTGTCTAAAATGTACAGATGT R: TGAAAAACATACCAACATTTCCGGAA	$(ATTCT)_{12}$	255-285	56.0	7	6.6176	0.7667	0.8633	0.8300	0.0043
HHF-65	KX548369	F: CGCAATGACAAAGGACAGGACAC R: TGTTGTGTAAGAGATGTTGATTGTTGCA	$(ATATG)_{10}$	230-240	58.0	3	2.9801	0.7667	0.6757	0.5904	0.4963
HHF-66	KX548370	F: CCATATGTCGTCGCACTAACCACT R: TGTGAACCCCTTTGCAGAATCTGT	$(CAAAA)_{25}$	239-264	57.3	6	5.3731	0.9333	0.8277	0.7869	0.7018
HHF-68	KX548371	F: AGTTGGCACCCCATACAACAAATA R: GATCATGCTGATCCAGTAATTTGTGAT	$(AGAAT)_{13}$	269-284	60.5	4	3.9046	0.7000	0.7565	0.6960	0.3299
HHF-77	KX548372	F: CAGTCACTCCACTATGCAGGAAGC R: ACATCCCCTGCACGACTTCTCTAA	$(TCTTC)_{16}$	232-267	58.0	5	4.6875	0.8667	0.8000	0.7527	0.4627
HHF-79	KX548373	F: TCTCAGCAAAAGTAAAACGGCAAC R: CGCTCGTGCTGGTCTGTAATAATT	$(TAGAA)_{11}$	222-247	56.0	6	5.7692	0.8000	0.8407	0.8024	0.5554

HHF-84	KX548374	F: ATGCAACCCTGTTACCCATTATGG R: GTCACTGGAATCTTATCACACCGC	$(TTATG)_{11}$	233-248	56.0	4	3.7267	0.8667	0.7441	0.6827	0.2651
HHF-88	KX548375	F: GCTTCCAACTGACAAAAGAAAGGT R: TGCCTTTACTTCTCACTTTTGATCAATT	$(TTTTG)_{11}$	220-255	59.5	5	4.5802	0.8000	0.7949	0.7460	0.6002
HHF-103	KX548376	F: GCCGTAGCATTTTTAGTCTTTTATCGT R: TGGGTATATAGGAGTGGTAATTGGGA	$(AAAAC)_{11}$	236-256	60.5	5	4.7619	0.7667	0.8034	0.7563	0.4657
HHF-104	KX548377	F: GCCCCTCTCTCTGATAGACAATGC R: AATGGTGACTGAGGCTGTCTTTCC	$(AAAGA)_{10}$	221-231	60.5	3	2.9851	0.6333	0.6763	0.5910	0.8223
HHF-108	KX548378	F: TTCATATGCAAGCGGAAACATCCA R: GCTGAAAGCTATGGTAGAAGACTTGA	$(TGTGT)_{14}$	200-250	60.0	11	9.6257	0.8667	0.9113	0.8868	0.6380
HHF-109	KX548379	F: GGAGCCCTTTGTTGAAACAGACTC R: TGAACAGCATTTCCTGACAAGAAGT	$(TATTC)_{11}$	254-274	57.1	5	4.8000	0.8667	0.8051	0.7582	0.2078
HHF-113	KX548380	F: CCAAGGGGGCTTCAAAACTGACTT R: TCCAGGTTTTGCAAGACCATTGGA	$(TTTTG)_{15}$	268-298	56.5	7	6.4286	0.8333	0.8588	0.8249	0.8443
HHF-120	KX548381	F: ACTGGATAACACTGGATGACCTTTAACA R: CCAATTCACCAGATCACCCAGAGG	$(TAACA)_{15}$	219-229	56.0	3	2.9557	0.5000	0.6729	0.5876	0.2342
HHF-123	KX548382	F: GGTTGATCTCATGTGAGTGACAGG R: AGCATTTTGTCTGTAGAATTGCTTAAGT	$(AAGAG)_{18}$	173-198	58.5	6	4.7493	0.6667	0.8028	0.7569	0.4288
HHF-148	KX548383	F: ACACTGTTTCCCTGTTAGCTAACCA R: TGCATTTTTATGTCCCATGGGCTT	$(GTAAC)_{26}$	192-217	59.2	6	5.9016	0.8667	0.8446	0.8069	0.0397
HHF-152	KX548384	F: GCGAACTGTGCGGTTGTTTATTCA R: AGCAAGGCAAGATGAGAAGAGACA	$(TTCTC)_{17}$	184-209	60.0	4	3.8217	0.7333	0.7508	0.6897	0.7432
HHF-153	KX548385	F: GCTGAATGGACTAAACACTCAAAAGCA R: CGACAGTCTGCCTGAATGATTCGA	$(AAAAC)_{10}$	233-253	59.5	5	4.7120	0.8667	0.8011	0.7538	0.9380
HHF-157	KX548386	F: TGGATACTTACAAGCAACTAAACCCT R: GCAGAACTAGTTAGTGAGACGCTG	$(AGAAG)_{13}$	195-205	57.0	3	2.9654	0.4333	0.6740	0.5888	0.0218
HHF-163	KX548387	F: GGCTGTTCTAAAACGCATGGCAAT R: CACATGTTCATTTCCCTTGAACAGCA	$(AAGAG)_{15}$	175-205	56.5	7	6.4516	0.8000	0.8593	0.8255	0.4931
HHF-177	KX548388	F: GGGGGTCTCTCTTGAAAGCGTTT R: TGTATTCTGCATCTTCATGTCGCAGA	$(TTTTG)_{17}$	146-156	60.5	3	2.8986	0.7333	0.6661	0.5812	0.8027

HHF-182	KX548389	F: CAGGTAGGTAGAACTCCAGGAACA R: AAAAAGCATACACAGATTCCAGGTTC	$(TTTCT)_{18}$	135-190	60.0	8	7.2874	0.9000	0.8774	0.8469	0.2576
HHF-184	KX548390	F: GTTAAGAGCCACATTTAACTTGAACAAC R: AAGAACTTTGGGCAACACATGTC	$(CTTTT)_{17}$	170-210	59.0	9	7.5630	0.8667	0.8825	0.8536	0.1741

Notes.

Size range, size range of fragment; bp, base pair; Tm, annealing temperature (℃); *Na*, number of alleles; *Ne*, effective number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; PIC, polymorphic information content; HWE, Hardy-Weinberg equilibrium.

Table 3(on next page)

Characteristics of 29 microsatellite markers in cross-species amplification for *M. amblycephala*, *M. hoffmanni*, *M. skolkovii* and *P. pekinensis*

Table 3 Characteristics of 29 microsatellite markers in cross-species amplification for *M. amblycephala*, *M. hoffmanni*, *M. skolkovii* and *P. pekinensis*

Locus	Species	Size range (bp)	Tm (℃)	*Na*	*Ne*	H_O	H_E	PIC	HWE *P* value
HHF-5	*M. amblycephala*	256	58.0	---	---	---	---	---	---
	M. hoffmanni	214	58.0	---	---	---	---	---	---
	M. skolkovii	201	56.0	---	---	---	---	---	---
	P. pekinensis	204	59.0	---	---	---	---	---	---
HHF-22	*M. amblycephala*	212	56.0	---	---	---	---	---	---
	M. hoffmanni	209	56.0	---	---	---	---	---	---
	M. skolkovii	193	58.5	---	---	---	---	---	---
	P. pekinensis	195	59.0	---	---	---	---	---	---
HHF-42	*M. amblycephala*	243	55.0	---	---	---	---	---	---
	M. hoffmanni	277	56.0	---	---	---	---	---	---
	M. skolkovii	330-335	55.0	2	1.9231	0.3333	0.4881	0.3648	0.0768
	P. pekinensis	365	55.0	---	---	---	---	---	---
HHF-58	*M. amblycephala*	241	57.5	---	---	---	---	---	---
	M. hoffmanni	196	57.5	---	---	---	---	---	---
	M. skolkovii	---	---	---	---	---	---	---	---
	P. pekinensis	---	---	---	---	---	---	---	---
HHF-61	*M. amblycephala*	261	56.5	---	---	---	---	---	---
	M. hoffmanni	---	---	---	---	---	---	---	---
	M. skolkovii	---	---	---	---	---	---	---	---
	P. pekinensis	---	---	---	---	---	---	---	---
HHF-62	*M. amblycephala*	234	59.3	---	---	---	---	---	---
	M. hoffmanni	222	56.0	---	---	---	---	---	---
	M. skolkovii	206-216	57.5	3	2.9079	0.6333	0.6672	0.5825	0.7221
	P. pekinensis	228-238	56.0	3	2.9557	0.6333	0.6729	0.5876	0.4160
HHF-63	*M. amblycephala*	275	55.6	---	---	---	---	---	---

	M. hoffmanni	268	57.4	---	---	---	---	---	---
	M. skolkovii	180	56.0	---	---	---	---	---	---
	P. pekinensis	---	---	---	---	---	---	---	---
	M. amblycephala	239	58.5	---	---	---	---	---	---
HHF-65	*M. hoffmanni*	243-248	59.3	2	1.9978	0.5667	0.5079	0.3747	0.5193
	M. skolkovii	219	58.4	---	---	---	---	---	---
	P. pekinensis	---	---	---	---	---	---	---	---
	M. amblycephala	225-235	59.8	3	2.9364	0.6667	0.6706	0.5851	0.6426
HHF-66	*M. hoffmanni*	196-206	60.5	3	2.9851	0.7333	0.6763	0.5910	0.7971
	M. skolkovii	268	57.4	---	---	---	---	---	---
	P. pekinensis	259	56.5	---	---	---	---	---	---
	M. amblycephala	315	58.5	---	---	---	---	---	---
HHF-68	*M. hoffmanni*	337	59.3	---	---	---	---	---	---
	M. skolkovii	316	60.0	---	---	---	---	---	---
	P. pekinensis	303	60.0	---	---	---	---	---	---
	M. amblycephala	336	56.0	---	---	---	---	---	---
HHF-77	*M. hoffmanni*	267	57.5	---	---	---	---	---	---
	M. skolkovii	286	56.0	---	---	---	---	---	---
	P. pekinensis	245	56.0	---	---	---	---	---	---
	M. amblycephala	232	58.5	---	---	---	---	---	---
HHF-79	*M. hoffmanni*	230-240	60.0	3	2.9801	0.7667	0.6757	0.5904	0.5422
	M. skolkovii	243	60.0	---	---	---	---	---	---
	P. pekinensis	241-266	60.0	6	5.6250	0.8667	0.8362	0.7971	0.7261
	M. amblycephala	198-223	59.5	4	3.9301	0.8333	0.7582	0.6981	0.3538
HHF-84	*M. hoffmanni*	290	58.0	---	---	---	---	---	---
	M. skolkovii	185	57.3	---	---	---	---	---	---
	P. pekinensis	---	---	---	---	---	---	---	---
	M. amblycephala	298	57.3	---	---	---	---	---	---
HHF-88	*M. hoffmanni*	224-239	60.5	3	2.9221	0.6000	0.6689	0.5837	0.0044
	M. skolkovii	273-283	60.0	3	2.7397	0.7333	0.6458	0.5594	0.6572

	P. pekinensis	315	60.0	---	---	---	---	---	---
HHF-103	*M. amblycephala*	190-195	58.5	2	1.9651	0.4667	0.4994	0.3705	0.7146
	M. hoffmanni	200-205	56.0	2	2.0000	0.4000	0.5085	0.3750	0.2347
	M. skolkovii	191	56.0	---	---	---	---	---	---
	P. pekinensis	191	59.0	---	---	---	---	---	---
HHF-104	*M. amblycephala*	---	---	---	---	---	---	---	---
	M. hoffmanni	---	---	---	---	---	---	---	---
	M. skolkovii	223	58.0	---	---	---	---	---	---
	P. pekinensis	200	58.0	---	---	---	---	---	---
HHF-108	*M. amblycephala*	162-167	58.0	2	1.9231	0.5333	0.4881	0.3648	0.6055
	M. hoffmanni	168	59.5	---	---	---	---	---	---
	M. skolkovii	161	59.5	---	---	---	---	---	---
	P. pekinensis	163	59.5	---	---	---	---	---	---
HHF-109	*M. amblycephala*	276	60.0	---	---	---	---	---	---
	M. hoffmanni	283	60.5	---	---	---	---	---	---
	M. skolkovii	256-266	60.0	3	2.9950	0.6000	0.6774	0.5920	0.1313
	P. pekinensis	257	59.5	---	---	---	---	---	---
HHF-113	*M. amblycephala*	294	56.0	---	---	---	---	---	---
	M. hoffmanni	---	---	---	---	---	---	---	---
	M. skolkovii	---	---	---	---	---	---	---	---
	P. pekinensis	---	---	---	---	---	---	---	---
HHF-120	*M. amblycephala*	194	56.5	---	---	---	---	---	---
	M. hoffmanni	199	59.0	---	---	---	---	---	---
	M. skolkovii	283	59.0	---	---	---	---	---	---
	P. pekinensis	282-302	58.5	5	4.4444	0.8000	0.7881	0.7390	0.9015
HHF-123	*M. amblycephala*	247	60.0	---	---	---	---	---	---
	M. hoffmanni	195-215	59.5	4	3.8217	0.8000	0.7508	0.6897	0.9423
	M. skolkovii	218-234	60.0	5	4.5000	0.7000	0.7910	0.7422	0.0425
	P. pekinensis	222	59.0	---	---	---	---	---	---
HHF-148	*M. amblycephala*	175	57.5	---	---	---	---	---	---

	M. hoffmanni	168	58.0	---	---	---	---	---	---
	M. skolkovii	---	---	---	---	---	---	---	---
	P. pekinensis	---	---	---	---	---	---	---	---
HHF-152	*M. amblycephala*	257	59.5	---	---	---	---	---	---
	M. hoffmanni	201	58.5	---	---	---	---	---	---
	M. skolkovii	220	60.0	---	---	---	---	---	---
	P. pekinensis	221	60.0	---	---	---	---	---	---
HHF-153	*M. amblycephala*	223	58.5	---	---	---	---	---	---
	M. hoffmanni	285	60.0	---	---	---	---	---	---
	M. skolkovii	226	56.0	---	---	---	---	---	---
	P. pekinensis	227	58.0	---	---	---	---	---	---
HHF-157	*M. amblycephala*	242	60.0	---	---	---	---	---	---
	M. hoffmanni	170	58.5	---	---	---	---	---	---
	M. skolkovii	248	60.0	---	---	---	---	---	---
	P. pekinensis	241	60.0	---	---	---	---	---	---
HHF-163	*M. amblycephala*	209-219	58.0	3	2.9950	0.6333	0.6774	0.5920	0.4231
	M. hoffmanni	---	---	---	---	---	---	---	---
	M. skolkovii	229	58.0	---	---	---	---	---	---
	P. pekinensis	238-248	56.0	3	2.8662	0.6333	0.6621	0.5760	0.6197
HHF-177	*M. amblycephala*	173-183	57.5	3	2.9851	0.6333	0.6763	0.5909	0.5694
	M. hoffmanni	181	59.3	---	---	---	---	---	---
	M. skolkovii	210	60.5	---	---	---	---	---	---
	P. pekinensis	208-218	60.0	3	2.8800	0.7000	0.6638	0.5786	0.3985
HHF-182	*M. amblycephala*	199	58.5	---	---	---	---	---	---
	M. hoffmanni	190	58.0	---	---	---	---	---	---
	M. skolkovii	211	60.5	---	---	---	---	---	---
	P. pekinensis	194	60.0	---	---	---	---	---	---
HHF-184	*M. amblycephala*	291	57.5	---	---	---	---	---	---
	M. hoffmanni	178	56.0	---	---	---	---	---	---
	M. skolkovii	231	58.0	---	---	---	---	---	---

	P. pekinensis	237	58.0	---	---	---	---	---	---

Notes.

Size range, size range of fragment; bp, base pair; Tm, annealing temperature (℃); *Na*, number of alleles; *Ne*, effective number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; PIC, polymorphic information content; HWE, Hardy-Weinberg equilibrium; ---, no value.