

Development and characterization of microsatellite markers for *Chaeturichthys stigmatias* (Actinopterygii: Gobiiformes: Gobiidae) based on restriction site-associated DNA sequencing (RAD-seq)

Bingjie CHEN^{1*}, Yu PAN^{1*}, Jian ZHENG¹, Chenyu SONG¹, Na SONG¹

¹ Key Laboratory of Mariculture (Ocean University of China), Ministry of Education, Qingdao, China

<https://zoobank.org/43A05001-59EB-4543-A974-2A921241C11D>

Corresponding author: Na Song (songna624@163.com)

Academic editor: Jolanta Kiełpińska ♦ **Received** 22 March 2022 ♦ **Accepted** 28 July 2022 ♦ **Published** 23 September 2022

Citation: Chen B, Pan Y, Zheng J, Song C, Song N (2022) Development and characterization of microsatellite markers for *Chaeturichthys stigmatias* (Actinopterygii: Gobiiformes: Gobiidae) based on restriction site-associated DNA sequencing (RAD-seq). Acta Ichthyologica et Piscatoria 52(3): 229–237. <https://doi.org/10.3897/aiep.52.84144>

Abstract

Chaeturichthys stigmatias Richardson, 1844, a fish species of the family Gobiidae, is an offshore warm-temperate fish species and a dominant component of estuarine ecosystems. In this study, restriction site-associated DNA sequencing was adopted to analyze the traits of candidate microsatellite markers for *C. stigmatias*, and 30 polymorphic loci were developed. A total of 5631 microsatellites with primer fragments were detected, among which trinucleotide repeats (57.56% of the total simple sequence repeats) were the most abundant, followed by di- (23.30%), tetra- (11.79%), penta- (4.14%), and hexa- (3.21%) nucleotide repeats type. The numbers of alleles per locus ranged from 6 to 14 with the mean value of 10.4. The mean value of observed heterozygosity and the expected heterozygosity were 0.349 and 0.870, respectively. The microsatellite locus with the lowest polymorphic information content (PIC) was 0.749, which indicated that all sites were highly polymorphic (PIC > 0.50). This is the first microsatellite development and characterization of this species to be reported.

Keywords

Chaeturichthys stigmatias, DNA sequences, microsatellite, polymorphic sites, primer detection, RAD-seq

Introduction

The branded goby, *Chaeturichthys stigmatias* Richardson, 1844 (also known as 矛尾鰕虎魚, finespot goby, or Cá Bống râu mắt nhỏ), is a warm-temperate nearshore benthic fish which is widely distributed in the coastal areas of China, Korea, and Japan (Sun et al. 2015). *Chaeturichthys stigmatias* expresses strong phenotypic

plasticity and can adapt to the changes of a variety of environmental factors, such as the bottom temperature and salinity (Liu et al. 2015). Although the important ecological values of *C. stigmatias* have been determined by biologists, the related studies mainly focused on the resource survey, community structure, feeding ecology, and fishery biological characteristics (Zhang et al. 2016; Meng et al. 2017; Li et al. 2018; Feng et al. 2019;

* These authors have contributed equally to this work and share first authorship.

Kindong et al. 2020; Han unpublished*). The information, however, about its genetic diversity and molecular markers is limited.

Exploring genetic diversity and population genetic structure may lead to a better understanding of the ecological importance of Gobiidae (see Yuan et al. 2012; Meng et al. 2017). Among molecular markers, microsatellites, or simple sequence repeats (SSR), is a simple repeat that is uniformly distributed in eukaryotic genomes and consists of tandem repeats of 2–6 nucleotides (Edwards et al. 1991). Given its codominance, high reproducibility, rich polymorphism, wide distribution, high stability, and easy detection (Dou et al. 2015; Song et al. 2016; Parthiban et al. 2018; Li et al. 2019), microsatellite markers have been used in a wide range of applications in population genetics, genetic breeding, evolutionary studies, identification of relations and individual identification (Lu et al. 2005; Hayes et al. 2007; Queirós et al. 2015), which can provide data reference and research guidance for them. In addition, microsatellite markers are greatly effective tools in population genetic studies because they could reveal the distinct population segments even in fine-scale genetic structure studies (Gandomkar et al. 2021). Therefore, it is essential to use microsatellite markers more conveniently and efficiently. However, the traditional methods for developing microsatellite markers are usually expensive, time-consuming, and cumbersome steps, with low coverage of loci in the genome, a long development cycle, and low versatility. Especially for non-model species with insufficient genetic information, the development of microsatellite markers is still difficult. In recent years, with the further maturity of the new generation of high-throughput sequencing technology and the rapid reduction of sequencing costs, a large number of plastid genomes, transcriptomes, and even genomes of non-model organisms have been sequenced, and the sequencing data in NCBI or other databases have increased significantly. As a reliable tool, high-throughput sequencing technologies optimize the field of discovery and development of molecular markers by generating large amounts of data (Shendure and Ji 2008; Stapley et al. 2010; Ekblom and Galindo 2011; Duan et al. 2017). In fact, high-throughput sequencing for developing SSR does not require sequencing depth as high as genome assembly and annotation, so the cost of this method is relatively low. Recently, high-throughput sequencing has been used to develop microsatellite markers in many fish, such as *Ctenopharyngodon idella* (Valenciennes, 1844) (see Yu et al. 2014), *Coilia nasus* Temminck et Schlegel, 1846 (see Fang et al. 2015), *Colossoma macropomum* (Cuvier, 1816) (see Ariede et al. 2018), *Genypterus chilensis* (Guichenot, 1848) (see González et al. 2019), and *Capoeta aculeata* (Valenciennes, 1844) (see Gandomkar et al. 2021).

Restriction site-associated DNA sequencing (RAD-seq) is a powerful tool to characterize the microsatellite

and single nucleotide polymorphism (SNP) markers, which was based on the second-generation sequencing technology (Khoshkholgh and Nazari 2020; Gandomkar et al. 2021). The reads generated by RAD-seq are grouped according to the enzyme recognition sequence, which could improve the precision and accuracy of contigs assembly, and improve the success rate of developing polymorphic microsatellite markers (Wei et al. 2014). In this study, RAD-seq was used to obtain preliminary data, and these data were applied to develop the *Chaeturichthys stigmatias* microsatellite primers, and finally, the validity of polymorphic primers was verified. The presently reported results may lay the foundation and provide references for the management and conservation of fishery resources.

Material and methods

Sampling and DNA extraction

A specimen of *Chaeturichthys stigmatias* was collected from the coast of Qingdao, China in November 2018, and sent for high-throughput sequencing. Dozens of *C. stigmatias* were collected from Zhoushan (August 2019), Qingdao (December 2019), Yantai (December 2019), and Weihai (October 2020), and 24 of them were used for polymorphism detection and genetic diversity analysis in this study. The samples were quickly dissected, and part of the muscle tissues on the caudal peduncle were collected and preserved in 95% alcohol in ice box, and then stored in -80°C for DNA extraction. The traditional phenol-chloroform method was used to extract genomic DNA. The total DNA was treated with RNase, and the DNA with high purity and without RNA contamination was obtained for the detection of SSR primers polymorphism. The extracted DNA was measured using a Nanodrop 2000 (Thermo Scientific, USA) and a Qubit 2.0 (Invitrogen, USA) bioanalyzer system.

RAD library construction and sequencing

After DNA quality inspection, library construction and sequencing were conducted. The steps of RAD library construction (Baird et al. 2008) were as follows: (1) Genomic DNA from the sample was digested at specific sites with a restriction enzyme, and the adapter P1 was ligated to the digested product. The P1 adapter contains forward amplification and Illumina sequencing primer sites, and an individual-specific nucleotide barcode; (2) The adapter-ligated fragments were then pooled, randomly sheared, and size-selected; (3) DNA was then ligated to a second adapter (P2), a Y adapter containing the reverse complement of the reverse amplification primer site, which en-

* Han DY (2013) [Study on feeding ecology of dominate gobiid fishes in Jiaozhou Bay.] Dissertation, Ocean University of China, Qingdao, China. [In Chinese]

sequences that lacking P1 adapter-ligated genomic fragments could not be amplified; (4) RAD tags with P1 adapter will be selected and amplified, and the 300–700 bp sequences were recovered. Agilent 2100 and Q-PCR were used to detect the size of library fragments and library quantification to determine whether the library meets the sequencing standard and then sequenced using the Illumina HiSeq2000 platform following the manufacturer's protocol. To obtain the clean reads, the reads with more than 10% N bases or low-quality bases ≤ 5 , adapter sequences, and duplicated sequences were discarded. The clean reads were used for subsequent analysis.

Detecting and verifying microsatellite primers

The software of detecting sequence repeats is “SSR search”, which is a Perl program written by Novogene (Beijing). The detection software is divided into three modules. The first module is used to detect all simple repeats of DNA sequence, the second module is to filter the results of the first module to remove the simple repeats that are too close. The detection criteria were as follows: the length of the SSR repeat unit ranges from 2 to 6 bp; the minimum length of the SSR sequence was 12 bp; the length of the upstream and downstream

sequences of the SSR was 100 bp, and 12 bp was the minimum distance between two SSR sequences. The third module is to use Primer3 (a software that designs primers under Linux or UNIX systems) to design primers (Rozen and Skaletsky 2000). The detected microsatellite primer sequences were further screened, and the screening criteria were as follows: the SSR units were repeated more than 6 times; the length of the SSR units ranges from 3 to 10 bp; the expected length of the PCR product was between 130 and 300 bp; the sequence of four consecutive bases was excluded. The selected primers were suitable for synthesizing SSR primers to verify primer polymorphism.

Microsatellites were verified through PCR and electrophoresis. Each 25 μ L PCR amplification system contained the following reagents: 17.25 μ L ultrapure water, 2.5 μ L $10 \times$ PCR buffer, 2 μ L dNTPs, 1 μ L each primer ($5 \mu\text{mol} \cdot \text{L}^{-1}$), 0.25 μ L Taq polymerase, and 1 μ L template DNA. The PCR reactions ran for 5 min at 94°C, followed by 38 cycles of 45 s at 94°C, 45 s at the annealing temperature (Table 1), and 45 s at 72°C in a thermal cycler. Cycling was followed by a final extension step at 72°C for 10 min. The PCR product was incubated at 4°C. The amplified PCR product was electrophoresed on an 8% non-denaturing polyacrylamide gel at 14 W for 3–4 h, and it was shown by silver staining (Lin et al. 2015). The allele size was identified according to the 20 bp DNA ladder.

Table 1. Characteristics of microsatellite loci in *Chaeturichthys stigmatias* from China.

Locus	Primer sequence (5'→3')		Repeat motif	T_a [°C]	Expected product length [bp]	N_a	H_o	H_e	PIC
	Forward	Reverse							
MW29	ACTAATTAGCATTTCAGCACCAGC	GTCATGCACAGTGACACCATAAT	(TG)15	58.3	135	7	0.000	0.823	0.778
MW31	TGATCGACAATGAAATGTAATG	TATTCTATAGCCACAGCTGCCT	(TG)7	56.4	145	13	0.292	0.895	0.866
MW32	TAAAGTGCCGTAACAAGTTGGAT	CGTCATGATTTCAGGAAGTAACA	(TA)10	55	144	9	0.042	0.870	0.836
MW34	AAGTGTCTATTCTGAGCGCACTT	TTGCAGTGATGAATCAAACATTC	(GAT)8	58.3	153	9	0.292	0.878	0.844
MW40	TCTGCATCTTCTGAACCTCACCT	CTCTGAAACACACAGTCACACCT	(GC)7	56.4	156	11	0.333	0.894	0.862
MW54	ATAGAAGGGACTTCAGTTGGACC	CCATTTAAACTCTGTCCAGACCCA	(AT)7	56.4	138	6	0.167	0.796	0.749
MW56	TGTATTCTCGCTTACTGCAGCTC	TCATTTCTCAGCATTGACTCTCAT	(ATA)7	56.4	132	14	0.375	0.883	0.854
MW66	AGAGTGAAAGAACGCAGTACC	GACCTTAGTGAGAGTGTGCGTGT	(CA)9	58.3	140	8	0.458	0.855	0.819
MW72	TGCAAACACTGCTTGTGTAGTT	TGAGCTGATTGTGTAGTTTGTCA	(TA)7	56.4	150	11	0.458	0.903	0.873
MW77	CTGCTGTGTTGTTACTCAGATG	TATCAAGGGCTCACTAAAGGACA	(GAG)7	58.3	137	6	0.250	0.821	0.774
MW79	GAAGAGGGAAGAGAGAACCAAAG	TTCTGTCCCAAATTCACCTCTC	(GA)9	56.4	160	10	0.417	0.834	0.793
MW80	TTAGACAGGACAGCGTTAGCATT	CACAGCAAAGGCTCTGAATACTT	(GA)7	56.4	147	13	0.417	0.867	0.834
MW83	GAGACACTGTCCAGAGCAGATCCT	TAATCAACAGCATGAAGAGCAGA	(GCT)7	56.4	148	10	0.500	0.840	0.801
MW86	AAATCCTTCTGCAATTGACTCTG	GAGAGGGAGGAAGAGATAATGGA	(CT)8	55	139	7	0.167	0.816	0.769
MW87	ACTGCTGTAGATTTACTGGTGC	TATCCTTATCCTCCTCTTTCACA	(TAC)8	60.2	157	11	0.417	0.876	0.842
MW88	TTGAGTATATTTAGCCCGTCTC	GCCGTTTGCTCATAACATAAACT	(AT)8	56.4	133	13	0.250	0.912	0.884
MW92	TTTGAAAAGGTGCAGGAGATG	TGAACTCCACTGCTCTGTGTA	(CT)15	53.1	136	12	0.708	0.908	0.879
MW97	CACAGCAAACAAGAAACAACAC	TATTACGGAAAGGGTAGGACCAT	(TAC)7	58.3	138	12	0.417	0.857	0.826
MW100	TCCCACCACAGAAGTTAAACAGT	GCATGTTCCCTTACAAGGTTTAC	(TAT)7	55	148	8	0.042	0.840	0.800
MW103	CTTTCTTACTTTCCCGCTCTCTT	CATGGAAATGGATAGAAATGGAA	(CT)7	55	133	11	0.375	0.889	0.858
MW104	AGGCAAGAAATATCACAGGGACT	TCGTGACTCATGGAATACCAAT	(AT)7	55	147	10	0.333	0.860	0.824
MW111	CAGGCCTGTTAGCTTAGCTGTAG	CACCTGGCACACACAACCTAAATA	(AT)12	58.3	139	12	0.542	0.893	0.862
MW113	GTATTTATCCGAGCACAGACTAC	TAAACGCACGAACAGTATCGTAA	(TG)12	55	156	12	0.833	0.898	0.868
MW115	TTATTTGCCAGTATTGACCCAGT	CCAAGCCTCTAAGAGTGTCTGAA	(CA)11	49.6	150	9	0.000	0.876	0.841
MW117	TGACGTGTGAACATTCGTGAGT	GAGGGAATGATGTCTGTGATTT	(ACA)8	58.3	151	13	0.417	0.931	0.905
MW118	TTATGGCCCTCAGTGTGTTAAT	CCTCGAGGAAATATCAGAGTATCG	(TAA)10	55	157	10	0.333	0.874	0.840
MW119	AAATGACGAGACAATTACAACCTGAT	TTCTTTGTGTATTATGGAAGTTCA	(TA)15	58.3	139	11	0.833	0.886	0.854
MW120	TTTCAGATACACTCATTGGACC	GAAACAACAGCAGTTGCACAAT	(AAT)7	60.2	140	13	0.500	0.897	0.867
MW121	TCTGTTTGTGACAGTGACAGAGT	CCTCCAGAGAAGGACTCATCAT	(TGC)7	58.3	132	12	0.167	0.903	0.873
MW123	TCCATCTAAACTGAACCAATG	TGAAATGTAGTCAATCTTTGCCA	(TTA)7	58.3	154	9	0.125	0.835	0.795

T_a = optimized annealing temperature, N_a = number of alleles, H_o = observed heterozygosity, H_e = expected heterozygosity, PIC = polymorphism information content.

Data analysis

After statistical analysis, the results were input into Genepop 4.0 (Rousset 2008). The parameters of SSR primers were calculated, including the mean value of effective allele number (N_e), polymorphism information content (PIC), observed heterozygosity (H_o) and expected heterozygosity (H_e), and Hardy–Weinberg equilibrium was also performed.

Ethics statement

We have read the policies relating to animal experiments and confirmed this study complied. All procedures performed in this study were approved by the Institutional Animal Care and Use Committee of the Ocean University of China.

Result and discussion

High-throughput sequencing and quality estimation.

A total of 4.682 Gb high-quality data was obtained, and the Q20 and Q30 values were 97.31% and 92.52%, respectively. The RAD-Tag capture rate was 98.03%, and the GC content was 39.43%. Genomic GC content had a significant effect on the randomness of second-generation genome sequencing. Too high (>65%) or too low (<25%) GC content will lead to sequencing bias and seriously affect the results of genomic analysis. The GC content of *Chaeturichthys stigmatias* was normal, and the sequencing quality was qualified, indicating that the sequencing of the database was successful (Zerbino and Birney 2008).

The sequences were clustered and assembled. The total contig base was 113 171 723 bp, and the total contig number was 337 800. The mean value of contig length of the assembly sequences was 335 bp, and N50 length was 393 bp. The GC content of the assembly result was 39.04%, which was consistent with the GC content of the sequencing clean data, indicating that the assembly result was true and reliable (Wang et al. 2017; Gao et al. 2018). Subsequently, the variation detection was carried out on

the assembly results. The number of heterozygous SNPs in the detected SNPs was 142 307, and the heterozygous rate was 82.47%. The high heterozygous SNP and the low homozygous SNP values also indicated the reliability of the assembly results.

Characterization of microsatellite loci

Based on the RAD-seq, the total number of identified microsatellites was 5829. Among them, there were 5631 microsatellite loci containing primer fragments (Table 2). The trinucleotide repeats were dominant (57.56%), followed by dinucleotide repeats (23.30%), tetranucleotide repeats (11.79%), pentanucleotide repeats (4.14%), and hexanucleotide repeats (3.21%).

The previous studies showed that the dominant repeating unit was discrepant. Some fish species were dinucleotide, such as *Megalobrama amblycephala* Yih, 1955 and *Larimichthys crocea* (Richardson, 1846) (see Wang et al. 2012; Zeng et al. 2013; Li unpublished*), while some fish were trinucleotide, such as *Acanthogobius hasta* (Temminck et Schlegel, 1845) and mollusk, *Ruditapes philippinarum* (see Yan et al. 2015; Song et al. 2019). The previous studies have suggested that enzymes and other proteins involved in various aspects of DNA processing and chromatin remodeling may be responsible for the taxonomic specificity of microsatellite abundance. This was manifested in that not only the repetitiveness of the genome varies, but also the dominant microsatellite types are different. This might indicate that SSRs play an important role in genome evolution, and the process responsible for the generation and fixation of SSR has also changed during evolution (Toth et al. 2000). In this study, trinucleotide repeats have absolute quantitative advantages, and the number of dinucleotide repeats was less than half. We speculate that a genetic mutation might occur during the evolution of *Chaeturichthys stigmatias*. Further comparative investigations including more species are needed to clarify this point.

The distribution and frequency of microsatellite motifs were presented in Fig. 1. The AT repeat motif (300) was the most frequent among all 11 types of dinucleotide repeat, whereas GC was the least frequent, with only one microsatellite locus. The AAT repeat motif (396) was the most frequent among all 60 types of a trinucleotide repeat. The AAAT repeat motif (55) was the most frequent among all 104 types of tetranucleotide repeat. The AATTG repeat motif (56) was the most frequent pentanucleotide repeat, and the ATTCTG (35) was the most frequent hexanucleotide repeat. Because the repeat types of trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide were too dispersed, only the top 30 types of loci were selected for illustration in order to show the results more clearly. All detailed data was provided in Suppl. material 1: Appendix 1 (online resource).

Table 2. Simple sequence repeat (SSR) distribution statistics for *Chaeturichthys stigmatias* from China.

Nucleotide repeat type	Statistics	
	SSR number	Percentage
Di-	1312	23.30
Tri-	3241	57.56
Tetra-	664	11.79
Penta-	233	4.14
Hexa-	181	3.21

* Li HM (2014) [New microsatellite satellite markers development based on whole genome sequencing information and its application in population genetics in large yellow croaker.] Dissertation, Zhejiang Ocean University, China. [In Chinese]

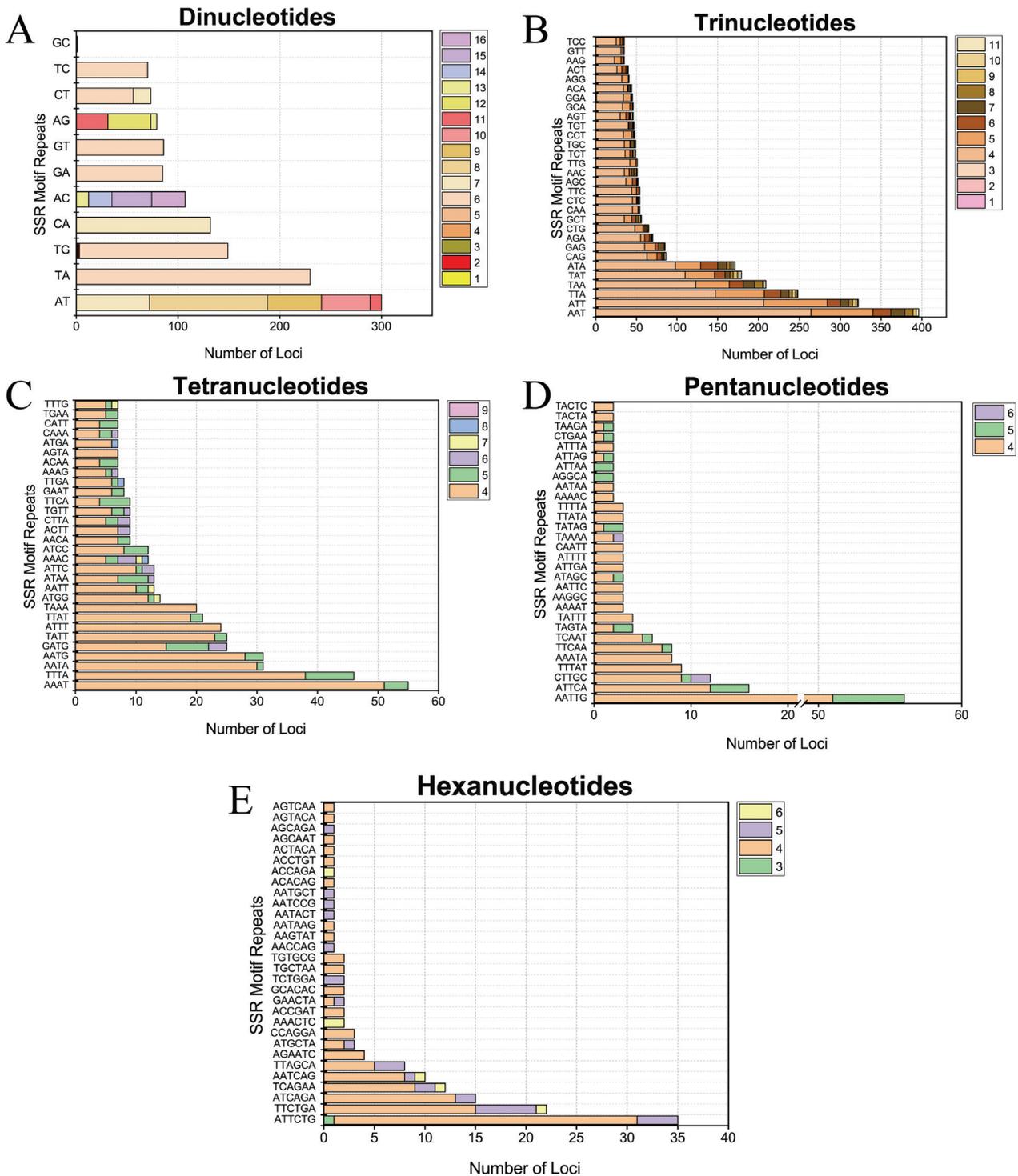


Figure 1. The distribution and frequency of microsatellite motifs of *Chaeturichthys stigmatias* from China. (A) Frequency of different dinucleotide microsatellite motifs; (B) Frequency of different trinucleotide microsatellite motifs; (C) Frequency of different tetranucleotide microsatellite motifs; (D) Frequency of different pentanucleotide microsatellite motifs; (E) Frequency of different hexanucleotide microsatellite motifs.

In terms of the frequency of repeating units, there were only four distinct types of repeats detected in pentanucleotide and hexanucleotide, and all of them were predominant at a frequency of a 4-fold repeat. Seven types were identified and 4-fold repeat was predominant in all tetranucleotide repeats. The types of repetition frequency detected in dinucleotide and trinucleotide were not less

than 10 types, and 5-fold repeat and 6-fold repeat were the main components in dinucleotide and trinucleotide respectively (Figs 1, 2).

In this study, the frequency distribution of the repetition units of dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide microsatellite were mainly 4–10 times, 4–7 times, 4–5 times, 4 times, and 4

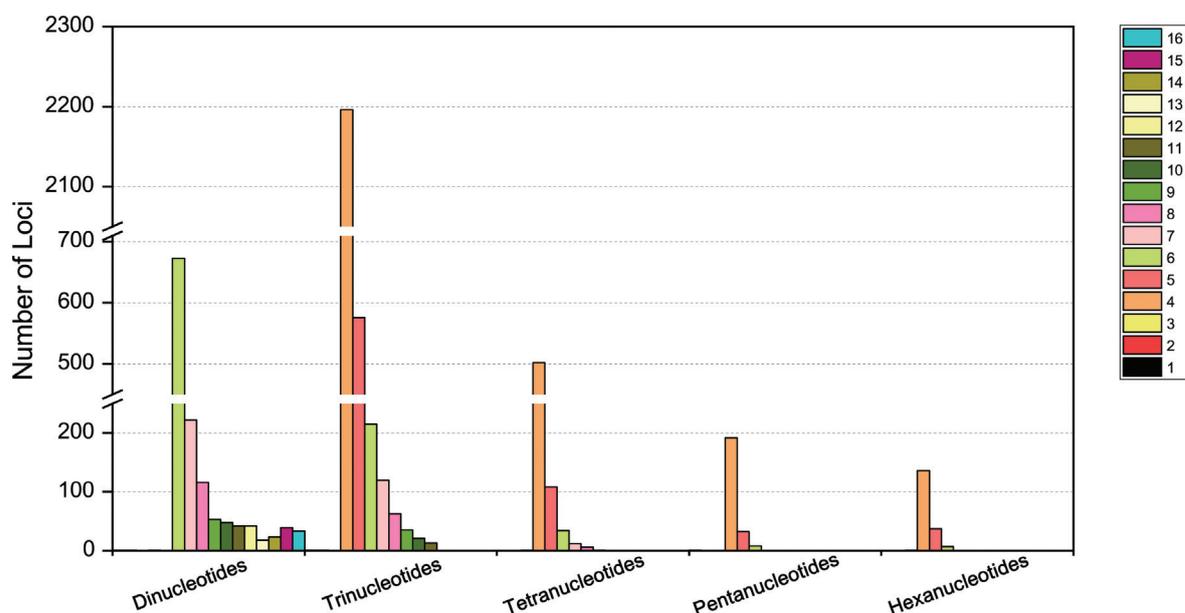


Figure 2. Frequency of repeating units of different types of microsatellites of *Chaeturichthys stigmatias* from China.

times, respectively (Fig. 2). This result showed that the frequency of tandem repeats decreased exponentially with the increase of repetition unit length, and this was consistent with the conclusion proposed by Chen et al. (2010) that the number of repeating units was negatively correlated with the length of repeating units. According to previous studies, slipped-strand mispairing is a major mechanism for DNA sequence evolution (Levinson and Gutman 1987), and the results of this study can be explained as microsatellites with a large number of repetitions may be more unstable due to the increase of sliding possibility (Ellegren 2004). It was generally believed that there was a certain positive correlation between the variation frequency of SSR sites and the number of repetition units (Schlötterer 2000). Katti et al. (2001) also reported that the mutation rate increases gradually with the increase of the length of repeating units in eukaryotes.

Detection of primer polymorphism

A gradient PCR experiment was performed on the synthesized 148 pairs of primers, and the optimal temperature of each pair of primers was screened. The results showed that a total of 97 pairs of primers were successfully amplified. Then after the PCR product was subjected to polyacrylamide gel electrophoresis experiments, a total of 30 primers with polymorphism were screened out. A total of 312 alleles were detected for 24 individuals at 30 polymorphic loci, and the number of alleles per locus ranged from 6 to 14, with the mean value of effective alleles was 10.4. The mean value of expected heterozygosity was 0.870, the observed heterozygosity was 0.349, and the mean value of polymorphic information content was 0.836 (Table 1). All the polymorphic sites deviated significantly from Hardy–Weinberg equilibrium ($P < 0.05$). The PIC was between

0.749 and 0.905, and all loci showed high polymorphism ($PIC > 0.5$) (Botstein et al. 1980).

In this study, 30 primers with polymorphism were screened out as dinucleotide and trinucleotide repeats, without tetranucleotide, pentanucleotide, and hexanucleotide repeats. Kong et al. (2019) observed that compared with trinucleotide and tetranucleotide repeats, dinucleotide repeats had a higher screening efficiency and polymorphism (Kong et al. 2019). However, in recent years, it has also been found that the trinucleotide and tetranucleotide repeats have higher screening efficiency and polymorphism than dinucleotide repeats in *Ctenopharyngodon idella*, *Hypophthalmichthys molitrix* (Valenciennes, 1844), and *Cyprinus carpio* Linnaeus, 1758 (see Fang et al. 2018). The higher repetition unit length has the disadvantages of lower repeats, lower sequence richness, and lower mutation rate. The better trinucleotide and tetranucleotide repeats polymorphisms obtained in other experiments may be related to the genome doubling in the long-term evolution of this species (Lu et al. 2009; Fang et al. 2018). The different results may be related to the specificity of species, the randomness of the number and type of primers selected in the experiment, and the number of samples. The above-mentioned results indicated that the SSR repeats which had higher screening efficiency and polymorphism may be species-dependent, and the most probable SSRs were dinucleotide and trinucleotide repeats. In terms of polymorphism, the mean PIC values of dinucleotide and trinucleotide repeats microsatellite primers screened in this study were 0.836 and 0.835, respectively, with little difference in polymorphism. Therefore, differences in screening efficiency and polymorphism may be caused by species differences or other factors (Kong et al. 2019).

The higher heterozygote ratio reflects the stability of the genetic structure of the population. We found that the observed heterozygosity (H_o) of 30 polymorphic sites was lower than the expected heterozygosity (H_e), show-

ing a relative lack of heterozygosity. It was generally believed that the loss of heterozygosity was caused by geographical isolation, decreased gene exchange between populations, and increased inbreeding (Zhao et al. 2009). The samples used in this study were collected from the Yellow Sea, the Bohai Sea, and the East China Sea. The low heterozygosity of *Chaeturichthys stigmatias* may be due to geographical isolation and excessive intraspecific hybridization. It was commonly accepted that the expected heterozygosity (H_e) was a more accurate reflection of the genetic diversity of a population than the observed heterozygosity (H_o) (Nei 1978). Therefore, the mean value of observed heterozygosity of 0.870 in this study showed a high population diversity.

At the same time, according to Hardy–Weinberg equilibrium analysis, all the 30 microsatellite loci discussed in this study showed significant imbalance, which was a common phenomenon in fish populations, such as *Siniperca scherzeri* Steindachner, 1892 and *Lutjanus peru* (Nichols et Murphy, 1922) (see Dou et al. 2015; Paz-García et al. 2017). This result also confirmed that these populations did not mate randomly, and non-random sampling was also the reason for the deviation of Hardy–Weinberg equilibrium. It was worth noting that inbreeding, subgroup structure, genetic drift, overfishing, Wallund effect, and ineffective alleles should also be considered (Bergh and Getz 1989; Lu et al. 2017; Song et al. 2018). The above results indicated

that the microsatellite markers identified in this study have high polymorphisms and can be used as effective molecular markers to analyze the genetic diversity and phylogenetic relations among *C. stigmatias*.

Conclusion

This study was conducted in combination with high-throughput sequencing, which also marks the first analysis of the microsatellite characteristics of *Chaeturichthys stigmatias*. In summary, a total of 4.682 Gb high-quality sequence data was obtained and 5631 SSRs were identified based on RAD-seq, indicating the high efficiency of the primer development of this technology. The 30 pairs of polymorphic primers obtained in this study will provide an effective basis for the future comparative analysis of the genetic structure and genetic characteristics of *C. stigmatias*, and also provide a significant basis for the development of microsatellite primers using high-throughput sequencing technology in the future.

Acknowledgments

We thank Dr Zonghang Zhang for the English editing. This work was supported by the National Key R and D Program of China (Grant number 2018YFD0900905).

References

- Ariede RB, Freitas MV, Hata ME, Matrochirico-Filho VA, Utsunomia R, Mendonça FF, Foresti F, Porto-Foresti F, Hashimoto DT (2018) Development of microsatellite markers using next-generation sequencing for the fish *Colossoma macropomum*. *Molecular Biology Reports* 45(1): 9–18. <https://doi.org/10.1007/s11033-017-4134-z>
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10): e3376. <https://doi.org/10.1371/journal.pone.0003376>
- Bergh MO, Getz WM (1989) Stability and harvesting of competing populations with genetic variation in life history strategy. *Theoretical Population Biology* 36(1): 77–124. [https://doi.org/10.1016/0040-5809\(89\)90024-5](https://doi.org/10.1016/0040-5809(89)90024-5)
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics* 32(3): 314.
- Chen M, Tan ZY, Zeng GM, Peng J (2010) Comprehensive analysis of simple sequence repeats in pre-miRNAs. *Molecular Biology and Evolution* 27(10): 2227–2232. <https://doi.org/10.1093/molbev/msq100>
- Dou YQ, Liang XF, Yang M, Tian CX, He S, Guo WJ (2015) Isolation and characterization of polymorphic EST-SSR and genomic SSR markers in spotted mandarin fish (*Siniperca scherzeri* Steindachne [sic]). *Genetics and Molecular Research* 14(4): 19317–19322. <https://doi.org/10.4238/2015.December.29.41>
- Duan X, Wang K, Su S, Tian R, Li Y, Chen M (2017) De novo transcriptome analysis and microsatellite marker development for population genetic study of a serious insect pest, *Rhopalosiphum padi* (L.) (Hemiptera: Aphididae). *PLoS One* 12(2): e0172513. <https://doi.org/10.1371/journal.pone.0172513>
- Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research* 19(6): 1349. <https://doi.org/10.1093/nar/19.6.1349>
- Eklom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107(1): 1–15. <https://doi.org/10.1038/hdy.2010.152>
- Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. *Nature Reviews, Genetics* 5(6): 435–445. <https://doi.org/10.1038/nrg1348>
- Fang DA, Zhou YF, Duan JR, Zhang MY, Xu DP, Liu K, Xu P, Wei Q (2015) Screening potential SSR markers of the anadromous fish *Coilia nasus* by *de novo* transcriptome analysis using Illumina sequencing. *Genetics and Molecular Research* 14(4): 14181–14188. <https://doi.org/10.4238/2015.November.13.1>
- Fang ZY, Chen XD, Wu YS, Tan JR, Zhang WE, Wang ZY, Lin TT, Zha GC, Shu H (2018) [Screening and characteristic analysis on Di-/Tri-/Tetra-nucleotide-repeated microsatellites in *Mastacembelus armatus*.] *Oceanologia et Limnologia Sinica* 49: 174–182. [In Chinese]
- Feng ZH, Zhang T, Li Y, He XR, Gao G (2019) The accumulation of microplastics in fish from an important fish farm and mariculture area, Haizhou Bay, China. *Science of the Total Environment* 696: 133948. <https://doi.org/10.1016/j.scitotenv.2019.133948>

- Gandomkar H, Shekarabi SPH, Abdolhay HA, Nazari S, Mehrgan MS (2021) Characterization of novel genotyping-by-sequencing (GBS)-based simple sequence repeats (SSRs) and their application for population genomics of *Capoeta aculeata* (Valenciennes, 1844). *Molecular Biology Reports* 48(9): 6471–6480. <https://doi.org/10.1007/s11033-021-06653-x>
- Gao SH, Yu HY, Wu SY, Wang S, Hu SN (2018) [Advances of sequencing and assembling technologies for complex genomes.] *Hereditas* (Beijing) 40(11): 944–963. [In Chinese] <https://doi.org/10.16288/j.ycz.18-255>
- González P, Dettliff P, Valenzuela C, Estrada JM, Valdés JA, Menezes C, Molina A (2019) Evaluating the genetic structure of wild and commercial red cusk-eel (*Genypterus chilensis*) populations through the development of novel microsatellite markers from a reference transcriptome. *Molecular Biology Reports* 46(6): 5875–5882. <https://doi.org/10.1007/s11033-019-05021-0>
- Hayes B, Baranski M, Goddard ME, Robinson N (2007) Optimisation of marker assisted selection for abalone breeding programs. *Aquaculture* 265(1–4): 61–69. <https://doi.org/10.1016/j.aquaculture.2007.02.016>
- Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* 18(7): 1161–1167. <https://doi.org/10.1093/oxfordjournals.molbev.a003903>
- Khoshkholgh M, Nazari S (2020) Characterization of single nucleotide polymorphism markers for the narrow-clawed crayfish *Pontastacus leptodactylus* (Eschscholtz, 1823) based on RAD sequencing. *Conservation Genetics Resources* 12(4): 549–553. <https://doi.org/10.1007/s12686-020-01154-8>
- Kindong K, Wu JH, Gao CX, Dai LB, Tian SQ, Dai XJ, Chen J (2020) Seasonal changes in fish diversity, density, biomass, and assemblage alongside environmental variables in the Yangtze River Estuary. *Environmental Science and Pollution Research International* 27(20): 25461–25474. <https://doi.org/10.1007/s11356-020-08674-8>
- Kong XL, Li M, Chen ZZ, Gong YY, Zhang P, Zhang J (2019) [Development and evaluation of di-/tri-nucleotide-repeated microsatellites by RAD-seq in *Decapterus macrosoma*.] *South China Fisheries Science* 15: 97–103. [In Chinese] <https://doi.org/10.12131/20180256>
- Levinson G, Gutman GA (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203–221. <https://doi.org/10.1093/oxfordjournals.molbev.a040442>
- Li ZY, Wu Q, Shan XJ, Yang T, Dai FQ, Jin XS (2018) [Keystone species of fish community structure in the Bohai Sea.] *Journal of Fishery Sciences of China* 25(2): 229–236. [In Chinese] <https://doi.org/10.3724/SP.J.1118.2018.17374>
- Li TT, Fang ZW, Peng H, Zhou JF, Liu PC, Wang YY, Zhu WH, Li L, Zhang QF, Chen LH, Li LL, Liu ZH, Zhang WX, Zhai WX, Lu L, Gao LF (2019) Application of high-throughput amplicon sequencing-based SSR genotyping in genetic background screening. *BMC Genomics* 20(1): 444. <https://doi.org/10.1186/s12864-019-5800-4>
- Lin L, Li CH, Chen ZZ, Xu SS, Liu Y (2015) Development and characterization of twenty-three microsatellite makers for the purpleback flying squid (*Symplectoteuthis oualaniensis*). *Conservation Genetics Resources* 7(1): 161–163. <https://doi.org/10.1007/s12686-014-0318-1>
- Liu X, Zhang CL, Ren YP, Xu BD (2015) [Spatiotemporal variation in the distribution and abundance of *Chaeturichthys stigmatias* in the Yellow River estuary and adjacent waters.] *Journal of Fishery Sciences of China* 22: 791–798. [In Chinese]
- Lu SQ, Liu Z, Liu HY, Xiao TY, Su JM (2005) [Microsatellite DNA analysis of genetic diversity and the phylogenetic relationships of four breed varieties of *Carassius* sp.] *Journal of Fishery Sciences of China* 12: 371–376. [In Chinese] <https://doi.org/10.3321/j.issn:1005-8737.2005.04.001>
- Lu CY, Mao RX, Li O, Geng LW, Sun XW, Ling LQ (2009) [Isolation and characterization of polymorphic tri- and tetranucleotide repeat microsatellite loci in common carp (*Cyprinus carpio*).] *Journal of Agricultural Biotechnology* 516(10): 3147–3151. [In Chinese]
- Lu YX, Di MY, Li YF, Zhou ZC, Hou HM, He CB, Wang S, Gao M (2017) [Microsatellite analysis of genetic diversity in wild and cultured populations of jellyfish *Rhopilema esculentum*.] *Fisheries Science* 36: 472–479. [In Chinese]
- Meng KK, Wang J, Zhang CL, Ren YP, Xu BD (2017) [The fishery biological characteristics of *Chaeturichthys stigmatias* in the Yellow River estuary and its adjacent waters.] *Journal of Fishery Sciences of China* 24(5): 939–945. [In Chinese] <https://doi.org/10.3724/SP.J.1118.2017.17083>
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89(3): 583–590. <https://doi.org/10.1093/genetics/89.3.583>
- 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: e144. <https://doi.org/10.1007/s13205-018-1172-8>
- Paz-García DA, Munguía-Vega A, Plomozo-Lugo T, Weaver AH (2017) Characterization of 32 microsatellite loci for the Pacific red snapper, *Lutjanus peru*, through next generation sequencing. *Molecular Ecology Resources* 44: 251–256. <https://doi.org/10.1007/s11033-017-4105-4>
- Queirós J, Godinho R, Lopes S, Gortazar C, de la Fuente J, Alves PC (2015) Effect of microsatellite selection on individual and population genetic inferences: An empirical study using cross-specific and species-specific amplifications. *Molecular Ecology Resources* 15(4): 747–760. <https://doi.org/10.1111/1755-0998.12349>
- Rousset F (2008) GENEPOP'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8(1): 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. In: Misener S, Krawetz SA (Eds) *Bioinformatics methods and protocols. Methods in Molecular Biology*. Vol. 132. Humana Press, Totowa, NJ, USA. <https://doi.org/10.1385/1-59259-192-2:365>
- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109(6): 365–371. <https://doi.org/10.1007/s004120000089>
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology* 26(10): 1135–1145. <https://doi.org/10.1038/nbt1486>
- Song N, Liu M, Yanagimoto T, Sakurai Y, Han ZQ, Gao TX (2016) Restricted gene flow for *Gadus macrocephalus* from Yellow Sea based on microsatellite markers: Geographic block of Tsushima Current. *International Journal of Molecular Sciences* 17(4): 467. <https://doi.org/10.3390/ijms17040467>

- Song N, Li PF, Zhang XM, Gao TX (2018) Changing phylogeographic pattern of *Fenneropenaeus chinensis* in the Yellow Sea and Bohai Sea inferred from microsatellite DNA: Implications for genetic management. *Fisheries Research* 200: 11–16. <https://doi.org/10.1016/j.fishres.2017.12.003>
- Song CY, Feng ZY, Li CH, Sun ZC, Gao TX, Song N, Liu L (2019) Profile and development of microsatellite primers for *Acanthogobius ommaturus* based on high-throughput sequencing technology. *Journal of Oceanology and Limnology* 38(6): 1880–1890. <https://doi.org/10.1007/s00343-019-9154-1>
- Stapley J, Reger J, Feulner PG, Smadja C, Galindo J, Ekblom R, Ben- nison C, Ball AD, Beckerman AP, Slate J (2010) Adaptation genomics: The next generation. *Trends in Ecology and Evolution* 25(12): 705–712. <https://doi.org/10.1016/j.tree.2010.09.002>
- Sun YN, Wei T, Jin XX (2015) Unusual features of control region and a novel NADH 6 genes in mitochondrial genome of the finespot goby, *Chaeturichthys stigmatias* (Perciformes, Gobiidae). *Mitochondrial DNA* 26(5): 665–667. <https://doi.org/10.3109/19401736.2013.840598>
- Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* 10(7): 967–981. <https://doi.org/10.1101/gr.10.7.967>
- Wang H, Zhang BW, Shi WB, Luo X, Zhou LZ, Han DM, Qing C (2012) [Structural characteristics of di-nucleotide/tetra-nucleotide repeat microsatellite DNA in *Pachyhynobius shangchengensis* genomes and its effect on isolation.] *Biodiversity Science* 20(1): 51–58. [In Chinese] <https://doi.org/10.3724/SP.J.1003.2012.08168>
- Wang JL, Zhu MX, Xu MH, Chen SL, Zhang FQ (2017) [Analysis on SSR in *Sinoswertia tetraptera* base on RAD-seq.] *Bulletin of Botanical Research* 37: 447–452. [In Chinese] <https://doi.org/10.7525/j.issn.1673-5102.2017.03.016>
- Wei N, Bemmels JB, Dick CW (2014) The effects of read length, quality and quantity on microsatellite discovery and primer development: From Illumina to PacBio. *Molecular Ecology Resources* 14(5): 953–965. <https://doi.org/10.1111/1755-0998.12245>
- Yan LL, Qin YJ, Yan XW, Wang LN, Bi CL, Zhang JY (2015) [Development of microsatellite markers in *Ruditapes philippinarum* using next-generation sequencing.] *Acta Hydrobiologica Sinica* 35: 1573–1580. [In Chinese] <https://doi.org/10.5846/stxb201305151071>
- Yu L, Bai J, Cao T, Cao TT, Fan JJ, Quan YC, Ma DM, Ye X (2014) Genetic variability and relationships among six grass carp *Ctenopharyngodon idella* populations in China estimated using EST-SNP markers. *Fisheries Science* 80: 475–481. <https://doi.org/10.1007/s12562-014-0709-y>
- Yuan YJ, Liu SF, Bai CC, Liu HB, Zhuang ZM (2012) Isolation and characterization of new 24 microsatellite DNA markers for golden cuttlefish (*Sepia esculenta*). *International Journal of Molecular Sciences* 13(1): 1154–1160. <https://doi.org/10.3390/ijms13011154>
- Zeng C, Gao ZX, Luo W, Liu XL, Wang WM (2013) [Characteristics of microsatellites in blunt snout bream (*Megalobrama amblycephala*) EST sequences using 454 FLX.] *Acta Hydrobiologica Sinica* 37(5): 982–988. [In Chinese] <https://doi.org/10.7541/2013.129>
- Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18(5): 821–829. <https://doi.org/10.1101/gr.074492.107>
- Zhang H, Xian WW, Liu SD (2016) Autumn ichthyoplankton assemblage in the Yangtze Estuary shaped by environmental factors. *PeerJ* 4: e1992. <https://doi.org/10.7717/peerj.1922>
- Zhao BL, Li ZB, Chen J, Lei GG, Zhang GL, Wang ZL (2009) [Heterozygosity in the nine populations of wild *Marsupenaeus Japonicus*.] *Journal of Jimei University(Natural Science)* 14: 21–26. [In Chinese] <https://doi.org/10.19715/j.jmuzr.2009.02.002>

Supplementary material 1

Appendix 1

Authors: Bingjie Chen, Yu Pan, Jian Zheng, Chenyu Song, Na Song

Data type: excel file

Explanation note: Detailed data for types of trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide loci.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/aiep.52.84144.suppl1>