Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers in swamp eel (Monopterus albus)

Hai-feng Tian, Qiao-mu Hu and Zhong Li
Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan, Hubei, China

Abstract

Objectives: Swamp eel is one model species for sexual reversion and an aquaculture fish in China. One local strain with deep yellow and big spots of Monopterus albus has been selected for consecutive selective breeding. The objectives of this study were characterizing the Simple Sequence Repeats (SSRs) of M. albus in the assembled genome obtained recently, and developing polymorphic SSRs for future breeding programs.

Methods: The genome wide SSRs were mined by using MISA software, and their types and genomic distribution patterns were investigated. Based on the available flanking sequences, primer pairs were batched developed, and Polymorphic SSRs were identified by using Polymorphic SSR Retrieval tool. The obtained polymorphic SSRs were validated by using e-PCR and capillary electrophoresis, then they were used to investigate genetic diversity of one breeding population.

Results: A total of 364,802 SSRs were identified in assembled M. albus genome. The total length, density and frequency of SSRs were 8,204,641 bp, 10,259 bp/Mb, and 456.16 loci/Mb, respectively. Mononucleotide repeats were predominant among SSRs (33.33%), and AC and AAT repeats were the most abundant di- and tri-nucleotide repeats motifs. A total of 287,189 primer pairs were designed, and a high-density physical map was constructed (359.11 markers per Mb). A total of 871 polymorphic SSRs were identified, and 38 SSRs of 101 randomly selected ones were validated by using e-PCR and capillary electrophoresis. Using these 38 polymorphic SSRs, 201 alleles were detected and genetic diversity level (Na, PIC, H\textsubscript{O}, and H\textsubscript{e}) was evaluated.

Corresponding author:
Zhong Li, Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, No.8, 1st Wudayuan Road, Donghu Hi-Tech Development Zone, Wuhan, 430223, China.
Email: lizhong@yfi.ac.cn

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Conclusions: The genome-wide SSRs and newly developed SSR markers will provide a useful tool for genetic mapping, diversity analysis studies in swamp eel in the future. The high level of genetic diversity ($Na = 5.29, PIC = 0.5068, H_O = 0.4665, H_e = 0.5525$) but excess of homozygotes ($F_{IS} = 0.155$) in one breeding population provide baseline information for future breeding program.

Keywords
Monopterus albus, swamp eel, simple sequence repeats, polymorphic SSRs, genetic diversity

Introduction
Simple sequence repeats (SSRs) or microsatellites are a group of DNA sequences consisting of tandemly repeated units (1–6 bp), and they are present in both protein coding and non-coding regions of the genome. Owing to the characteristics of simplicity, abundance, ubiquity, co-dominance, and highly polymorphism, SSRs have become a common tool in fish and aquaculture, including genetics, genomics, characterization of genetic stocks, marker-assisted selection, parentage determination, and linkage analysis and quantitative trait loci (QTL) mapping. Recently, genome-wide identification and development of SSR markers have been successfully performed in various fish, such as spotted sea bass (Lateolabrax maculatus), common carp, and spotted scat (Scatophagus argus), and many marine animals.

Swamp eel (Monopterus albus) has been a model species for studying sexual reversal and it is becoming an important freshwater aquaculture fish in China and South-East Asia due to its flavor and nutrition. As different genetic lineages and multiple local strains of swamp eel are present in China, one local strain with deep yellow and big spots that mainly distributed in Jianghan plain has been selected for consecutive selective breeding due to superiority in growth rate and fecundity. Marker-assisted selection (MAS) would provide great help for accelerating the breeding process in fish, and SSRs have been extensively applied in such breeding programs in many aquaculture fishes. However, only limited SSRs had been developed in swamp eel. Although genome-wide SSRs have been characterized swamp eel, only a total of 99,293 loci were identified, and few of them have been validated in this species.

Here, based on the genome assembly of the deep yellow and big spots strain of swamp eel in our previous work, we characterized the density, type, and distribution of SSRs motifs, and developed polymorphic genomic SSR markers in this work. This study contributes to increasing the number of molecular markers available for genetic studies in M. albus, which will allow the development of breeding programs in the future.

Materials and methods
Animal care
The experimental procedure was approved by the Animal Experimental Ethical Inspection of Laboratory Animal Centre, Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences (ID Number: 2020-THF-01).
Identification of genome wide SSRs

The reference genome sequences of *M. albus* (with deep yellow and big spots) were downloaded from Genome Sequence Archive in National Genomics Data Center under accession number CRA003062. Perl scripts from MISA were used to perform SSR identification with the default parameters (http://pgrc.ipk-gaterslen.de/misa/). Initially SSRs of 1–6 nucleotides motifs were identified with the minimum repeat unit defined as 10 for mononucleotides, 6 for dinucleotides, 5 for trinucleotides, 4 for tetra-nucleotides, and 3 each for penta- and hexa-nucleotides. Compound SSRs were defined as those with an interval less than 100-nt between two repeat motifs. For simplicity, those repeats with unit patterns being circular permutations and reverse complements were grouped as one type. For instance, AGG denotes AGG, GGA, GAG, CCT, CTC, and TCC in different reading frames or on the complementary strand. Relative frequency (SSRs per megabase pair (Mbp)) and relative density (SSRs in base pairs (bp) per Mb) were used to help conduct comparisons between different repeat types.

Identified SSRs were assigned to genomic compartments using the BEDtools intersect tool v2.29.0. The sequences and coordinates of gene models, exons, coding sequences (CDSs), and intronic and intergenic regions for the swamp eel genomes were determined according to the positions in the genome annotation files in GFF format reported previously.

Primer design for genome wide SSRs

To design primers for identified SSRs, two interface Perl script modules (p3_in.pl and p3_out.pl) were used to interchange data between MISA and the primer designing software Primer 3. The primer design criteria were as follows: primer length was between 18 and 27 bp with 20 bp as optimum, melting temperature was from 57°C to 63°C, GC content ranged from 30% to 70%, and product size was from 100 to 280 bp. An R package (http://www.R-project.org), Cmplot, was used to draw the high-density physical map of the newly developed SSR markers.

Screen and validation of polymorphic SSRs

Based on our assembled genome and the available transcriptomes of swamp eel, Polymorphic SSR retrieval tool (PSR) were deployed to detect SSR length polymorphisms of perfect repeats. Allele number was taken as a measurement to evaluate the polymorphism of each type of motif among different samples.

We only choose polymorphic SSRs with tri- or more nucleotide repeat motifs to validate in the following analysis. Initially, primers were designed and validated their amplification efficiency by using electronic PCR (e-PCR) over our sequenced genome data as templates. Then a total of 101 newly developed primer pairs for SSRs were randomly selected and tested for polymorphism with six individuals by capillary electrophoresis of amplicons. The 5’ ends of the forward primer of each primer pair was labeled with FAM (6-carboxyfluorescein). The
PCR reactions were conducted on Veriti 384 well PCR System (Applied Biosystems). PCR amplification was carried out in a total volume of 10 μL containing 2 × Taq PCR MasterMix (Gene Tech), 0.25 μM primer (each direction), and about 20 ng genomic DNA as the template. The genomic DNA of *M. albus* was extracted from tail muscle using the magnetic beads extraction Kit (NanoMagBio, China). DNA concentration was measured using a Nanodrop (ThermoFisher). A two-stage touchdown amplification program was used. An initial denaturation step of 5 min at 95°C was used, followed by a first stage in which the annealing temperature was gradually decreased from 62°C to 52°C through 10 cycles. This treatment was followed by a second stage consisting of 25 amplification cycles, in which the annealing temperature was set 52°C. Denaturation and elongation steps were held constant for 30 s at 95°C and 30 s at 72°C, respectively, during both stages. An elongation step of 20 min at 72°C was performed after stage 2. PCR products were diluted with ddH2O 10× (PCR product of multiplex mix I) and 20× (PCR product of multiplex mix II). Each diluted PCR product was then mixed with 8 μL of Hi-Di™ Formamide mixed with 0.5 μL of GeneScan 500 LIZ Size Standard (Thermo Fisher Scientific). Fragment lengths were determined using GeneMarker 2.2 (SoftGenetics, State College, USA) after capillary gel electrophoresing on an ABI 3730xl Genetic Analyzer (Life Technologies, Carlsbad, CA, USA).

The genotypic data were initially processed in Microsoft Office Excel 2007, manually checked for errors, and then transformed into input files that were required for subsequent analyses. The number of alleles (*N*ₐ), the effective number of alleles (*N*ₑ), observed heterozygosity (*H*₀), expected heterozygosity (*H*ₑ), Shannon’s Information Index, and Fixation index (*F*) were calculated for each locus using GENALEX 6.5.35 The polymorphism information content (PIC) was implemented by Excel-microsatellite-toolkit version 3.1.36 Deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were tested using the exact tests of Genepop version 4.7.5 (https://kimura.univ-montp2.fr/~rousset/Genepop.htm) with 10,000 dememorization and in 200 batches with 5000 iterations per batch.37 The *p* values for HWE and LD were corrected for multiple comparisons by applying a sequential Bonferroni correction. MICROCHECKER v.2.2.3 was used to test for the possibility of scoring errors, allelic dropout, and null alleles.

**Results**

**Identification of SSRs in the *M. albus* genome**

The available 799.72 Mb genome sequences of *M. albus* were searched for the presence of different SSRs. A total of 364,802 SSRs were identified, and compound types (43,851) constituted 12.02% of the total SSRs. The total lengths of the identified SSRs were 8,204,641 bp, accounting for 1.03% of the whole genome length. The frequency and density of SSRs were 456 loci/Mb and 10,259.35 bp/Mb, respectively. Of 287,884 perfect SSRs (Table 1 and Figure 1), mononucleotide repeats
were represented in maximum counts (121,595) followed by di- (93,972), tri- (19,499), tetra- (18,675), penta- (26,757), and hexa-nucleotides (7386).

We detected 413 types of SSR motifs in total (Supplemental Table 1). There were 2, 4, 10, 32, 100, 265 types of mono- to hexanucleotide repeats in *M. albus*, respectively. Among different types of repeats, the motif A had the highest occurrence (37.63%) in the eel genome, followed by AC (24.31%), AG (4.87%), C (4.6%), AT (3.4%) (Supplemental Table 1). The most frequent mono- to hexa-nucleotides motifs were A (89.09%), AC (74.47%), AAT (37.65%), ATCC (24.61%), AAATC (25.57%), and ACACAG (22.22%) (Figure 2). Among dinucleotide repeats, the motif AC had the highest occurrence (74.47%), followed by AG (14.92%) and AT (10.53%), and CG repeats were the lowest in proportion (0.07%). Amongst the 10 types of trinucleotide repeats identified, AAT repeats were the most abundant (37.65%), followed by AGG (12.99%), AAG (11.44%), AAC (10.77%), AGC (9.82%), ATC (9.51%), ACC (7.83%), and the other three repeats (ACT, ACG, and CCG) were uncommon (together representing 3.91%). For tetra-, penta-, and hexanucleotide repeats, the most abundant motifs were ATCC (24.61%), AAATC (25.57%), and ACACAG (22.22%), respectively.

We then analyzed the distributions of SSRs on each chromosome of the swamp eel genome. A total of 316,657 SSR markers were distributed on 12 chromosomes and the number of SSRs on the chromosomes ranged from 18,962 to 35,822 (Table 2). The average frequency of SSRs was 404.32 SSRs per Mb, ranging from 381.49 SSR per Mb on chromosome 11 to 407.47 SSR per Mb on chromosome 8 (Table 2). Noticeably, when mononucleotide SSRs was not considered, the highest frequency of SSR markers was found on chromosome 3 and the least one was on Chromosome 12 (Table 2). Besides, about 21% of the identified SSRs were present in unanchored scaffolds and contigs (data not shown). The percentage of mononucleotide repeats was the highest, followed by di-, and pentanucleotide repeats. The percentage of tri- and tetranucleotides were nearly equivalent, and the percentage of hexanucleotide repeats was the lowest (Supplemental Figure 1).

The distribution of SSRs in different genomic regions, including exons, 5’UTR, 3’UTR, CDSs, intronic, and intergenic regions, were determined. The occurrences

| Nucleotide     | Total counts | Total length (bp) | Average length (bp) | Frequency (loci/Mb) | Density (bp/Mb) | Percent of SSRs in the whole genome (%) |
|----------------|--------------|-------------------|---------------------|---------------------|-----------------|----------------------------------------|
| Mononucleotide | 121,595      | 1,579,653         | 12.99               | 152.05              | 1975.26         | 33.33                                  |
| Dinucleotide   | 93,972       | 2,149,414         | 22.87               | 117.51              | 2687.71         | 25.76                                  |
| Trinucleotide  | 19,499       | 369,276           | 18.94               | 24.38               | 461.76          | 5.35                                   |
| Tetranucleotide| 18,675       | 359,552           | 19.25               | 23.35               | 449.60          | 5.12                                   |
| Pentanucleotide| 26,757       | 453,095           | 16.93               | 33.46               | 566.57          | 7.33                                   |
| Hexanucleotide | 7386         | 156,438           | 21.18               | 9.24                | 195.62          | 2.02                                   |

**Table 1.** The number, length, frequency, and density of six different types of perfect SSRs.
and relative frequencies of SSRs were found to differ significantly in coding and noncoding regions (Figure 3(a)). SSRs were most commonly located in intergenic regions, followed in order by intronic regions, exons, and CDSs (Figure 3(b)). Noticeably, the frequencies of SSRs in 3’UTRs (13.27 SSRs/Mb) are higher than those in 5’UTRs (2.52 SSRs/Mb) and CDSs (2.87 SSRs/Mb) (Figure 3(b)).

In CDSs, trinucleotide SSRs were the most abundant type, followed by hexa-, di-, penta-, mono-, and tetra-nucleotide SSRs (Figure 4(a)). In exons, mononucleotide SSRs were the most abundant type, followed by di-, tri-, penta-, tetra-, and hexa-nucleotide SSRs (Figure 4(b)). In intronic and intergenic regions, mononucleotide SSRs were the most abundant type, followed in decreasing order by di-, penta-, tetra-, and tri-nucleotide SSRs, and hexanucleotide SSRs were the least abundant type (Figure 4(c) and (d)).

**Development of the genome-wide SSR primers**

The flanking sequences of all identified SSRs were used to design suitable forward and reverse primer pairs. A total of 284,775 SSR markers were successfully developed on 12 chromosomes in swamp eel, which accounted for 78.02% of all identified SSRs. Based on the start positions of SSR markers, we successfully anchored these markers to the reference genome physical map, and there is no great difference of marker density among chromosomes on the physical map (Figure 5). This high-density physical map would provide an opportunity for accelerating mapping and breeding applications of different traits.
Table 2. Number of SSRs on the 12 chromosome of *M. albus*.

| Chr. | SSR marker | Mono | Di  | Tri | Tetra | Penta | Hexa | Compound | Chr. length (Mb) | Frequency* (loci/Mb) |
|------|------------|------|-----|-----|-------|-------|------|----------|------------------|---------------------|
| Chr1 | 35,822 | 14,150 | 10,412 | 2097 | 2036 | 2795 | 571 | 3761 | 88.66 | 404.04 (244.44) |
| Chr2 | 35,129 | 13,635 | 10,241 | 2046 | 2132 | 2795 | 585 | 3695 | 87.52 | 401.38 (245.59) |
| Chr3 | 30,402 | 11,031 | 9561 | 1836 | 1679 | 2551 | 543 | 3201 | 75.26 | 403.96 (257.39) |
| Chr4 | 28,368 | 10,975 | 8296 | 1730 | 1675 | 2367 | 427 | 2898 | 70.93 | 399.96 (245.22) |
| Chr5 | 27,705 | 10,561 | 8173 | 1667 | 1533 | 2518 | 448 | 2805 | 69.38 | 399.35 (247.12) |
| Chr6 | 27,170 | 10,334 | 7922 | 1660 | 1616 | 2224 | 473 | 2941 | 67.24 | 404.07 (250.38) |
| Chr7 | 24,710 | 9266 | 7173 | 1594 | 1529 | 2147 | 466 | 2535 | 63.37 | 389.94 (243.72) |
| Chr8 | 25,575 | 9922 | 7282 | 1491 | 1557 | 2217 | 416 | 2690 | 62.77 | 407.47 (249.39) |
| Chr9 | 22,363 | 8649 | 6443 | 1488 | 1314 | 1836 | 452 | 2181 | 57.70 | 387.55 (237.66) |
| Chr10 | 20,861 | 7867 | 6308 | 1263 | 1235 | 1732 | 351 | 2105 | 52.27 | 399.12 (248.61) |
| Chr11 | 18,962 | 7018 | 5518 | 1255 | 1187 | 1741 | 385 | 1858 | 49.70 | 381.49 (240.30) |
| Chr12 | 19,590 | 7999 | 5291 | 1274 | 1144 | 1576 | 338 | 1968 | 49.01 | 399.74 (236.52) |

*SSR frequency excluding mononucleotide was given in parenthesis.*
Figure 2. The distribution of the major repeat types in the genome of *M. albus*. 
Screen and validation of polymorphic SSRs

The sequenced genome and reported transcriptome provided opportunities for identifying the repeat size variations rapidly. As genotyping mononucleotide SSRs can be error prone, we only selected from di- to hexanucleotide repeats (162,612 SSRs) to screen polymorphic SSRs by using PSR tool, and 871 SSRs showed polymorphisms across those 12 individuals among those perfect SSRs (Supplemental Table 2).

Among 101 SSR loci tested, 63 loci were excluded due to PCR failure, monomorphism, unexpected product size, or multiple peak profiles in the following genotyping analysis. The other 38 ones were validated to be polymorphic among six individuals and used for genotype analysis furtherly (Table 3).

Genetic variations at 38 loci were assessed in a panel of 29 individuals of our breeding population. The number of alleles per locus ($N_A$), number of effective alleles per locus ($N_E$), Shannon’s information index ($I$), observed ($H_O$) and expected ($H_e$) heterozygosity, unbiased expected heterozygosity ($uH_e$), fixation index ($F$), and polymorphic information content (PIC) are presented in Table 3. The PIC values ranged from 0.066 to 0.896 (0.507 in average). Twenty SSR markers were highly informative (PIC $> 0.5$), and 14 ones were in medium polymorphism degree ($0.5 > $PIC$ > 0.25$). A total of 201 alleles were identified in 38 loci. $N_A$ varied from 2

![Figure 3. Comparisons of percentage (a) and frequency (b) of SSRs in different genomic regions of M. albus.](image-url)
The Shannon information index (I) ranged from 0.174 to 2.432, with a mean of 1.09. \( H_O \) and \( H_E \) ranged from 0.069 to 0.828 and from 0.067 to 0.904, respectively. The fixation index (\( F \)) ranged from 0.091 to 0.477 with an average of 0.152, and 19 loci presented heterozygote deficiencies (\( F_{IS} > 0.15 \)). Only two of the 38 loci (Ma30 and Ma50) showed significant deviations from HWE after sequential Holm-Bonferroni corrections (\( p < 0.05 \)). Besides, only one significant deviate from linkage equilibrium (loci Ma65 and Ma71) (following the sequential Holm-Bonferroni correction, \( p < 0.05 \)) of the 703 pairwise comparisons in the tested population, and none of the remaining markers showed significant LD (\( p > 0.06 \)).

**Discussion**

In this study, a total of 364,802 SSRs were identified based on the complete genome of *M. albus*. Even only considering from di- to hexanucleotide repeats, the
identified number (223,012) and frequency (278 loci/Mb) of SSRs in this study is much larger than the previous reports obtained by using restriction site associated DNA (RAD) sequencing (9897 loci, 195 loci/Mb) or second generation genome sequencing (99,293 loci, 179 loci/Mb). Considering the criteria selected for identifying the SSRs in our work are the same as those used previously, such great number of SSRs identified here might be attribute to the high continuity of the assembled genome based on PacBio and Hi-C sequencing. The total length of all SSRs from *M. albus* genome was 8.2 Mb, and positive correlations between the genome size and total length of SSRs ($R^2 = 0.773$, $p < 0.01$) were also confirmed when compared with those marine fish reported recently, suggesting that this rule might also work in teleost fishes.

The motif A was the most abundant mononucleotide motif in swamp eel, which is also found in different mosquitoes and some marine animals. The most abundant repeat motifs of di- and trinucleotide repeats seem to be conserved. AC was the most abundant repeat categories in dinucleotides and CG was the least one, which are consistent with the cases in most vertebrates reported by Jiang et al. and in other eukaryotes. AAT was the predominant trinucleotides type, which is similar to the situation in human, camelid, and many teleost fishes. Interestingly, the higher proportions of penta-nucleotide than tri- and tetranucleotide motifs are consistent with the result characterized by using RAD sequencing, but different from those reported by Li et al., and the proportions of each repeat motifs of tri-, tetra-, penta-, and hexanucleotide repeats are different from those

![Figure 5. Overview of the high-density SSR physical map in swamp eel (*M. albus*). The bar represents the number of SSR markers within a 1-Mb window.](image)
Table 3. Characteristics of 38 SSR loci developed for 29 samples of *M. albus*.

| Locus name | Repeat motif | Forward primer | Reverse primer | Allele size range (bp) | N | Na | Ne | I | Ho | He | uHe | F | PIC | P of HWE |
|------------|--------------|----------------|----------------|------------------------|---|----|----|---|----|-----|-----|----|-----|-------|
| Ma1        | (ATT)9       | F: TGCTTAAGGAATCTCACGGC | R: CAGGATGCAAGGTTGGTGCTA | 219–228 | 29 | 4 | 2.292 | 0.935 | 0.379 | 0.564 | 0.574 | 0.327 | 0.467 | 0.0163 |
| Ma3        | (CTG)6       | F: GAGCTTCTCTGGTCTCTGATG | R: TGCCCTTGTACTGGATAGCCC | 195–201 | 29 | 3 | 1.324 | 0.479 | 0.138 | 0.245 | 0.249 | 0.437 | 0.227 | 0.0068 |
| Ma4        | (GCT)7       | F: TAGGGCAGTGTAGGATGTTG | R: TTGGTGTCTTCAGTCAGGG | 202–220 | 29 | 6 | 1.682 | 0.874 | 0.345 | 0.405 | 0.413 | 0.15 | 0.383 | 0.1522 |
| Ma5        | (TCTAC)7     | F: TTCAGGTTCGGGACTCTTGAC | R: AGACGGTTGTTCAGAGGAGG | 228–248 | 29 | 5 | 3.174 | 1.321 | 0.621 | 0.685 | 0.697 | 0.094 | 0.632 | 0.7913 |
| Ma6        | (TGA)5       | F: TGTTGAAAGTGTGATGATGA | R: ACAGCAGGTATACCTGGTGC | 138–144 | 29 | 3 | 1.576 | 0.653 | 0.241 | 0.366 | 0.372 | 0.34 | 0.326 | 0.0238 |
| Ma23       | (CAG)8       | F: TGACCCCACTACAGCAAGA | R: TCTTGCTTTCGTCACACAGG | 194–200 | 29 | 3 | 1.275 | 0.418 | 0.172 | 0.216 | 0.22 | 0.201 | 0.199 | 0.3364 |
| Ma24       | (CTG)9       | F: TCTGGTCTGTCAGAGGAGAT | R: AGCTGCAGCAGGAGGAATAC | 121–136 | 29 | 6 | 3.324 | 1.349 | 0.759 | 0.699 | 0.711 | -0.085 | 0.646 | 0.548 |
| Ma26       | (AGA)8       | F: TGTTCCACCTCGACACAGAG | R: TGGTGAAACACCAAGAGAACC | 179–194 | 29 | 6 | 3.609 | 1.488 | 0.448 | 0.723 | 0.736 | 0.38 | 0.688 | 0.0058 |
| Ma28       | (ATC)7       | F: AGAGGGACGTGACTGGCAGCT | R: GTGAACCCCTGAAACAGACGG | 134–167 | 29 | 9 | 2.307 | 1.307 | 0.483 | 0.567 | 0.577 | 0.148 | 0.545 | 0.0191 |
| Ma29       | (CTT)10      | F: GGAGAGTACGGTGTCGTTCGA | R: GCAGAAGATGCTTCAAGAGG | 239–254 | 29 | 5 | 1.97 | 0.927 | 0.379 | 0.492 | 0.501 | 0.229 | 0.441 | 0.2016 |
| Ma30       | (GAT)7       | F: GCACCTGGGACCTTTCAGACT | R: GGGTTTCTAAGCTTGGATAGG | 259–271 | 29 | 5 | 3.746 | 1.439 | 0.483 | 0.733 | 0.746 | 0.341 | 0.691 | 0.0005* (0.019) |
| Ma32       | (GCT)6       | F: CTTGGCATTGGAAGAAGCTCCA | R: GAAACACCTGGCGTCTCCTC | 236–251 | 29 | 6 | 3.754 | 1.444 | 0.483 | 0.734 | 0.747 | 0.342 | 0.689 | 0.0023 |
| Ma39       | (GAT)8       | F: AGGTGAAGGGAACACAGCTGT | R: TTGCCCTCTGTCTGACCTCT | 190–226 | 29 | 11 | 5.357 | 1.971 | 0.793 | 0.813 | 0.828 | 0.025 | 0.794 | 0.9374 |
| Ma41       | (TAGT)5      | F: TGAAGCAGGATTTAACGAGGG | R: AAAGAAGGCTTTTCTTCTT | 183–187 | 29 | 2 | 1.859 | 0.655 | 0.241 | 0.462 | 0.47 | 0.477 | 0.355 | 0.0139 |

(continued)
| Locus name | Repeat motif | Forward primer | Allele size range (bp) | N | Na | Ne | I | Ho | He | uHe | F | PIC | P of HWE |
|------------|--------------|----------------|-----------------------|---|----|----|---|----|----|-----|---|-----|-----------|
| Ma42 (TCT)12 | R:GACAAGGTTGAGTTCGGCTC F:ATCCTGGAGAGGGGTGTTTTG R:CGTCCCCAAAAGAAAGTGAA | 214–241 | 29 | 8 | 4.765 | 1.721 | 0.828 | 0.79 | 0.804 | -0.047 | 0.759 | 0.495 |
| Ma48 (TTGG)6 | F:GGCATATTGGCTGACAGTT R:GGGTCGACTAAACACCCAGC | 223–243 | 29 | 6 | 1.842 | 1 | 0.483 | 0.457 | 0.465 | -0.056 | 0.438 | 0.8148 |
| Ma49 (CAG)10 | F:AGATTCAGAGGGGAAGGCC | 212–224 | 29 | 5 | 2.609 | 0.915 | 0.552 | 0.517 | 0.526 | -0.068 | 0.441 | 0.4112 |
| Ma50 (GTA)5 | F:TTGCCGTATGTTTCTGAAA R:AACTCACCAAGTTTGGACGC | 274–283 | 29 | 4 | 3.081 | 1.163 | 0.483 | 0.675 | 0.687 | 0.285 | 0.606 | 0.0006* (0.022) |
| Ma53 (TGA)5 | F:ATGATTCAGTGAAACGCCACA R:CCACCTTTTTTGCTCAGGT | 112–121 | 28 | 3 | 1.736 | 0.752 | 0.393 | 0.424 | 0.432 | 0.074 | 0.381 | 0.3531 |
| Ma54 (ATC)8 | F:GTACAGCAGCCCCGGAAGTACG R:CGGACTTCGAGCAGCATTTGA | 155–167 | 29 | 5 | 2.666 | 1.219 | 0.621 | 0.625 | 0.636 | 0.007 | 0.580 | 0.554 |
| Ma61 (CAT)10 | F:GGCCGGCTTGCAGGAAAAATTA R:TCAAAATGAGGGAAAAACCG | 183–225 | 29 | 13 | 10.447 | 2.432 | 0.828 | 0.904 | 0.92 | 0.085 | 0.896 | 0.1461 |
| Ma62 (GAG)6 | F:GAGCGAGAAGAACGATGAGG R:TGACAGAAGAGCTCCAGACT | 182–191 | 29 | 4 | 1.901 | 0.864 | 0.379 | 0.474 | 0.482 | 0.199 | 0.425 | 0.235 |
| Ma65 (AAAC)5 | F:CCACTGTAGTGTGTGAGCGAC R:TACACAGCGTTGGTGAAAGAT | 213–241 | 29 | 6 | 2.616 | 1.238 | 0.483 | 0.618 | 0.629 | 0.218 | 0.569 | 0.1886 |
| Ma66 (AAC)8 | F:GTTTGCGCTCTTGGGCACCTT R:ACCTTTTTGCACACACTTTGCAC | 168–174 | 29 | 3 | 1.072 | 0.174 | 0.069 | 0.067 | 0.068 | -0.027 | 0.066 | 0.1 |
| Ma68 (CCT)6 | F:TTGAACCAACAGGTTCCCAACA R:ACTTACTTGCCCAAGGAGGTTG | 138–144 | 29 | 3 | 1.644 | 0.635 | 0.414 | 0.392 | 0.399 | -0.056 | 0.327 | 0.3561 |
| Ma71 (TCA)13 | F:TCACAAGAGCCAGAGAAGC R:GGCGAGCATCAAGACACCTCTC | 254–269 | 29 | 6 | 2.993 | 1.271 | 0.586 | 0.666 | 0.678 | 0.12 | 0.607 | 0.4916 |
| Ma73 (CAG)7 | F:ACATTTTCTTGGTCTGGAAACG R:TCTAGGCTTGGTTCCTTTCCTC | 247–256 | 29 | 4 | 2.717 | 1.136 | 0.69 | 0.632 | 0.643 | -0.091 | 0.569 | 0.9701 |

(continued)
| Locus name | Repeat motif | Forward primer | Reverse primer | Allele size range (bp) | N  | Na | Ne | I  | Ho | He | uHe | F   | PIC | P of HWE |
|-----------|--------------|----------------|----------------|-----------------------|----|----|----|----|----|----|-----|-----|-----|---------|
| Ma74      | (CTT)10      | F:GGGATTGTCTGGAATGCTGT | R:CGGAAACCAAAAAGTGGAAGA | 107–128 | 28  | 8  | 4.115 | 1.631 | 0.607 | 0.757 | 0.771 | 0.198 | 0.720 | 0.153 |
| Ma77      | (GTG)7       | F:CCTCTTTGGATAGTGTTGGTTG | R:CTGCAACAATGGGAGTTTCTCT | 112–124 | 29  | 4  | 1.488 | 0.663 | 0.276 | 0.328 | 0.334 | 0.159 | 0.308 | 0.193 |
| Ma78      | (TAAGA)5     | F:CTGACGGTTGTTGTGTAACGG | R:CACTGACGGCGTTAAAACCTT | 201–231 | 29  | 6  | 3.285 | 1.349 | 0.448 | 0.696 | 0.708 | 0.356 | 0.641 | 0.010 |
| Ma79      | (TCA)7       | F:CCACAACCTCCTCCTCTCTC | R:ACTGAGTGGTTTGTCTGACGC | 262–301 | 29  | 11 | 3.098 | 1.624 | 0.586 | 0.677 | 0.689 | 0.134 | 0.655 | 0.163 |
| Ma80      | (AGC)7       | F:CTCTTTCCGTCACTGGAAG | R:ATGGCTGAATTTGCTCTTGT | 106–118 | 29  | 4  | 1.439 | 0.647 | 0.241 | 0.305 | 0.31  | 0.209 | 0.291 | 0.091 |
| Ma85      | (TGC)9       | F:GTGGAAGAAGCTGGATGAGC | R:CACATGGGCTGTCTCACTG | 204–228 | 29  | 7  | 3.834 | 1.547 | 0.571 | 0.739 | 0.753 | 0.227 | 0.699 | 0.022 |
| Ma96      | (TCA)7       | F:GAACATATCCAGCCATCCCAA | R:GGACAGCGGAAGGACTCTG | 153–159 | 28  | 3  | 1.992 | 0.751 | 0.5  | 0.498 | 0.507 | −0.004 | 0.391 | 0.655 |
| Ma97      | (TCG)5       | F:CAAAGCTGAACCCAAACCTC | R:ATGACTCATGACGAGCAGA | 124–130 | 29  | 3  | 2.67  | 1.03  | 0.552 | 0.625 | 0.636 | 0.118 | 0.547 | 0.424 |
| Ma98      | (TTG)6       | F:AAACCAACAGCCAAACCAA | R:AAAAAGGGGAAAACATCTG | 200–206 | 29  | 3  | 2.288 | 0.914 | 0.586 | 0.563 | 0.573 | −0.041 | 0.474 | 0.207 |
| Ma99      | (AGAA)5      | F:CACTGCTTCATGAGGAACAT | R:TCTCCGCTGCTCTTTTCTTT | 160–176 | 29  | 5  | 1.286 | 0.522 | 0.241 | 0.222 | 0.226 | −0.086 | 0.215 | 1 |
| Ma101     | (AAG)7       | F:AGTTTACACAGCCACAAAAC | R:CCCAAATCCGCTTCTTCTT | 139–148 | 29  | 3  | 2.789 | 1.06  | 0.345 | 0.641 | 0.653 | 0.462 | 0.567 | 0.048 |

F: fixation index; (He-Ho)/He; He: expected heterozygosity; Ho: observed heterozygosity; I: Shannon's information index; N: number of samples; Na: number of different alleles; Ne: number of effective alleles; PIC: polymorphism information content; P of HW: Hardy-Weinberg probability test; uHe: unbiased expected heterozygosity.

*Indicated deviation from Hardy-Weinberg equilibrium (p < 0.05) after Bonferroni's correction.
reported in two recent works about swamp eel. Whether such discrepancy could be explained by using different sets of germplasm needed to be evaluated in future.

SSRs are powerful markers for population genetic analysis, QTL mapping, and other related genetic and genomic studies. Here, 284,775 SSR markers were developed successfully, and a highly density physical map was constructed, which might be used for future marker assisted breeding and QTL mapping. A total of 871 polymorphic SSRs were also obtained through comparing sequence variants among 12 public available transcriptomes, and 38 loci were validated by capillary gel electrophoresing. As SSRs with PIC more than 0.5 and $H_e$ more than 0.6 are the most reasonable informative loci for application in genetics, 20 SSRs are highly informative and would be very useful genetic markers for future genetic diversity, population genetic studies, and molecular breeding programs. Altogether, there was a high level of genetic diversity (Na = 5.3, $H_o$ = 0.469, $H_e$ = 0.554, PIC = 0.508) for the population studied here, which is similar to that of one wild population found in Jiangxi province (Na = 5.7, $H_o$ = 0.3795, $H_e$ = 0.5397, PIC = 0.4987), but relatively lower than four wild populations found in Hubei, Anhui, Jiangsu, and Zhejiang provinces along the middle and lower reaches of the Yangtze River Basin.

Considering the SSRs developed in this study were mainly trinucleotide repeats, which are typically associated with a low level of variability, the allele number ($N_a$) and PIC (Table 3) suggest that the genetic diversity within this broodstock is not low in contrast to those mainly using dinucleotide repeats. Noticeably, the average $H_o$ (0.467) was slightly less than the $H_e$ (0.553), and the mean value of the fixation index ($F$) was 0.152, indicating that a deficiency of heterozygotes and inbreeding within this breeding population.

**Conclusion**

We identified 364,802 SSRs which belong to 413 types of motifs, and developed 284,775 potential SSR markers accordingly. A total of 871 polymorphic SSRs were identified and 38 ones were validated. These newly identified SSRs and developed high variable SSR markers would enrich SSR marker resources and provide a resourceful dataset for genetic, genomic, and evolutionary biology studies of this species along with SSR markers from previous reports. Using 38 validated polymorphic SSR loci, the high level of genetic diversity and existence of inbreeding were revealed in our breeding population, which would provide a background of knowledge for future breeding programs.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, author-ship, and/or publication of this article: This work was supported by the Central Public-Interest Scientific Institution Basal Research Fund, Chinese Academy of Fishery Sciences (CASF) (Grant numbers 2020XT08 and 2020TD33).

ORCID iD
Hai-feng Tian https://orcid.org/0000-0002-8264-5579

Supplemental material
Supplemental material for this article is available online.

References
1. Chistiakov DA, Hellemans B and Volckaert FAM. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* 2006; 255: 1–29.
2. Liu ZJ and Cordes JF. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 2004; 238: 1–37.
3. Sundaray JK, Rasal KD, Chakrapani V, et al. Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. *Aquacult Int* 2016; 24: 1089–1102.
4. da Silva BC, Pereira A, Massago H, et al. Genetic characterization of selected *Nile tilapia* in Santa Catarina. *Semin Cienc Agrar* 2020; 41: 1739–1754.
5. Mastrochirico-Filho VA, del Pazo F, Hata ME, et al. Assessing genetic diversity for a pre-breeding program in piaractus mesopotamicus by SNPs and SSRs. *Genes* 2019; 10: 668.
6. Zhang J, Ma WG, Wang WM, et al. Parentage determination of yellow catfish (*Pelteobagrus Fulvidraco*) based on microsatellite DNA markers. *Aquacult Int* 2016; 24: 567–576.
7. Guo W, Tong J, Yu X, et al. A second generation genetic linkage map for silver carp (*Hypophthalmichys molitrix*) using microsatellite markers. *Aquaculture* 2013; 412: 97–106.
8. Fan S, Huang H, Liu Y, et al. Genome-wide identification of microsatellite and development of polymorphic SSR markers for spotted sea bass (*Lateolabrax maculatus*). *Aquac Rep* 2021; 20: 100677.
9. Ji P, Zhang Y, Li C, et al. High throughput mining and characterization of microsatellites from common carp genome. *Int J Mol Sci* 2012; 13: 9798–9807.
10. Huang Y, Jiang D, Li M, et al. Genome survey of male and female spotted scat (*Scatophagus argus*). *Animals (Basel)* 2019; 9(12): 1117.
11. Jiang Q, Li Q, Yu H, et al. Genome-wide analysis of simple sequence repeats in marine animals-a comparative approach. *Mar Biotechnol* 2014; 16: 604–619.
12. Zhang Y, Zhang S, Liu ZX, et al. Epigenetic modifications during sex change repress gonadotropin stimulation of Cyp19a1a in a teleost ricefield eel (*Monopterus albus*). *Endocrinology* 2013; 154: 2881–2890.
13. Zhang YZ, Sun X, Zhang LH, et al. Testicular Dnmt3 expression and global DNA methylation are down-regulated by gonadotropin releasing hormones in the ricefield eel *Monopterus albus*. *Scientific Reports* 2017; 7(1): 1–13.

14. Liu CK. Rudimentary hermaphroditism in Symbranchoid eel, *Monopterus javanensis*. *Sinensia* 1944; 15: 1–8.

15. Yang D, Chen F and Ruan G. Aquaculture of the Paddy Eel, *Monopterus albus*. In: Gui JF, Tang Q, Li Z, et al. (eds) *Aquaculture in China: success stories and modern trends*. London: John Wiley & Sons, 2018, pp.283–296.

16. Khanh N and Ngan H. Current practices of rice field eel *Monopterus albus* (zuiw, 1973) culture in Vietnam. *Aquaculture Asia Mag* 2010; 15: 26–29.

17. Silva AA, Suay ES and Silva WS. Notes on freshwater clam and land snail as alternative food for rice paddy eel, *Monopterus albus*. *Int J Fish Aquat Stud* 2020; 8(3): 223–226.

18. Liang HW, Guo SS, Li Z, et al. Assessment of genetic diversity and population structure of swamp eel Monopterus albus in China. *Biochem Syst Ecol* 2016; 68: 81–87.

19. Cai X, Yu SM, Mipam T, et al. Phylogenetic lineages of *Monopterus albus* (Synbranchiformes: Synbranchidae) in China inferred from mitochondrial control region. *J Zool Syst Evol Res* 2013; 51: 38–44.

20. Martinez V, Guimarães E, Ruane J, et al. (eds) *Marker-assisted selection in fish and shellfish breeding schemes. Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish*. Rome, Italy: Food and Agriculture Organization of the United Nations, 2007, pp.329–362.

21. Yue GH. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish Fish* 2014; 15: 376–396.

22. Lei L, Feng L, Jian TR, et al. Characterization and multiplex genotyping of novel microsatellites from Asian swamp eel, *Monopterus albus*. *Conserv Genet Resour* 2012; 4: 363–365.

23. Li WT, Liao XL, Yu XM, et al. Isolation and characterization of polymorphic microsatellites in a sex-reversal fish, rice field eel (*Monopterus albus*). *Mol Ecol Notes* 2007; 7: 705–707.

24. Liu Z, Luo X, Lu S, et al. Screening of microsatellite and genetic polymorphism of Monopterus albus populations with different gender phenotype. *Jiangsu nong ye xue bao* 2009; 25: 333–338.

25. Zhang F, Wang Y, Xu M, et al. Development and characterization of twenty-eight polymorphic microsatellite for rice field eel (*Monopterus albus*) using RAD tag sequencing. *Genet Mol Res* 2018; 17: gmr16039880.

26. Li Z, Chen F, Huang C, et al. Genome-wide mapping and characterization of microsatellites in the swamp eel genome. *Sci Rep* 2017; 7: 3157.

27. Tian H-F, Hu Q-M and Li Z. A high-quality de novo genome assembly of one swamp eel (*Monopterus albus*) strain with PacBio and Hi-C sequencing data. *G3 (Bethesda)* 2021; 11(1): jkaa032.

28. Beier S, Thiel T, Munch T, et al. MISA-web: a web server for microsatellite prediction. *Bioinformatics* 2017; 33: 2583–2585.

29. Jurka J and Pethiyagoda C. Simple repetitive DNA sequences from primates: compilation and analysis. *J Mol Evol* 1995; 40: 120–126.

30. Karaoğlu H, Lee CMY and Meyer W. Survey of simple sequence repeats in completed fungal genomes. *Mol Biol Evol* 2005; 22: 639-649. DOI: 10.1093/molbev/msi057.

31. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; 26: 841–842.
32. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res* 2012; 40(15): e115.
33. Ruan R, Chen M, Yue HM, et al. Comparative transcriptome analysis of livers from three strains of Chinese swamp eels. *Aquac Res* 2020; 51: 5251–5258.
34. Cantarella C and D’Agostino N. PSR: polymorphic SSR retrieval. *BMC Res Notes* 2015; 8: 525.
35. Peakall R and Smouse PE. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 2012; 28: 2537–2539.
36. Shaibi T, Lattorff HMG and Moritz RFA. A microsatellite DNA toolkit for studying population structure in Apis mellifera. *Mol Ecol Resour* 2008; 8: 1034–1036.
37. Rousset F. GENEPOP ’007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 2008; 8: 103–106.
38. Van Oosterhout C, Hutchinson WF, Wills DPM, et al. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 2004; 4: 535–538.
39. Wang XT, Zhang YJ, Qiao L, et al. Comparative analyses of simple sequence repeats (SSRs) in 23 mosquito species genomes: identification, characterization and distribution (Diptera: Culicidae). *Insect Sci* 2019; 26: 607–619.
40. Toth G, Gaspari Z and Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 2000; 10: 967–981.
41. Subramanian S, Mishra RK and Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol* 2003; 4: R13.
42. Manee MM, Algarni AT, Alharbi SN, et al. Genome-wide characterization and analysis of microsatellite sequences in camelid species. *Mammal Res* 2020; 65: 359–373.
43. Vieira MLC, Santini L, Diniz AL, et al. Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol* 2016; 39: 312–328.
44. Botstein D, White RL, Skolnick M, et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980; 32(3): 314.
45. Zhou Y, Hu H, Zhang L, et al. Microsatellite analysis of genetic diversity of *Monopterus albus* along the middle and lower reaches of Yangtze River Basin. *Biotechnol Bull* 2011; 11: 187–192. (In Chinese)
46. Cho YG, Ishii T, Temnykh S, et al. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 2000; 100: 713–722.
47. Gadissa F, Tesfaye K, Dagne K, et al. Genetic diversity and population structure analyses of *Plectranthus edulis* (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. *BMC Genet* 2018; 19: 92.

**Author biographies**

Hai-feng Tian is an Assistant Professor at Yangtze River Fisheries Research Institute. Currently, he mainly focused on the development of genetic markers to investigating population diversity and marker assisted breeding of swamp eels (*Monopterus albus*).
Qiao-mu Hu is an Associate Professor at Yangtze River Fisheries Research Institute. He is mainly interested in the sexual determination mechanisms of fishes and amphibia, and now working on the mechanisms of sexual reversion of swamp eel (Monopterus albus).

Zhong Li is a Professor at Yangtze River Fisheries. He mainly focused on genetic breeding of freshwater fishes, and had successfully bred a novel allotetraploid gibel carp (Carassius auratus gibelio) and a novel silver carp (Hypophthalmichthys molitrix). Currently, he is mainly working on the selective breeding of swamp eel.