Characterization and Development of EST-SSR Markers Derived from Transcriptome of Yellow Catfish

Jin Zhang 1, Wenge Ma 1, Xiaomin Song 1, Qiaohong Lin 1, Jian-Fang Gui 1,2,* and Jie Mei 1,*

1 Key Laboratory of Freshwater Animal Breeding, Ministry of Agriculture, Freshwater Aquaculture Collaborative Innovation Center of Hubei Province, College of Fisheries, Huazhong Agricultural University, Wuhan 430070, China
2 State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Wuhan 430072, China

* Author to whom correspondence should be addressed; E-Mails: jfgui@ihb.ac.cn (J.-F.G.); jmei@mail.hzau.edu.cn (J.M.); Tel./Fax: +86-27-6878-0707 (J.-F.G.); +86-27-8728-2113 (J.M.).

Received: 6 August 2014; in revised form: 28 September 2014 / Accepted: 29 September 2014 / Published: 13 October 2014

Abstract: Yellow catfish (Pelteobagrus fulvidraco) is one of the most important freshwater fish due to its delicious flesh and high nutritional value. However, lack of sufficient simple sequence repeat (SSR) markers has hampered the progress of genetic selection breeding and molecular research for yellow catfish. To this end, we aimed to develop and characterize polymorphic expressed sequence tag (EST)–SSRs from the 454 pyrosequencing transcriptome of yellow catfish. Totally, 82,794 potential EST-SSR markers were identified and distributed in the coding and non-coding regions. Di-nucleotide (53,933) is the most abundant motif type, and AC/GT, AAT/ATT, AAAT/ATTT are respective the most frequent di-, tri-, tetra-nucleotide repeats. We designed primer pairs for all of the identified EST-SSRs and randomly selected 300 of these pairs for further validation. Finally, 263 primer pairs were successfully amplified and 57 primer pairs were found to be consistently polymorphic when four populations of 48 individuals were tested. The number of alleles for the 57 loci ranged from 2 to 17, with an average of 8.23. The observed heterozygosity (Ho), expected heterozygosity (He), polymorphism information content (PIC) and fixation index (FIS) values ranged from 0.04 to 1.00, 0.12 to 0.92, 0.12 to 0.91 and −0.83 to 0.93,
respectively. These EST-SSR markers generated in this study could greatly facilitate future studies of genetic diversity and molecular breeding in yellow catfish.

**Keywords:** EST-SSRs; yellow catfish; 454 pyrosequencing; genetic diversity

1. Introduction

Molecular marker systems, such as simple sequence repeats (SSRs) or microsatellites [1], single nucleotide polymorphism (SNPs) [2], amplified fragment length polymorphisms (AFLPs) [3] and random amplification of polymorphic DNAs (RAPDs) [4] have been developed and are applied to fisheries and aquaculture. Yellow catfish is an important freshwater fish for its delicious flesh and high market value, whereas overfishing is decreasing its number and genetic diversity [5]. Applying genomic tools in the selection of elite broodstock has the potential to improve the productivity and commercial value of this species. In populations of yellow catfish, males grow faster than females by two to three folds. For this reason, an all-male monosex population has been massively produced for commercial purpose [3,6,7]. However, genetic resources and suitable molecular markers are still scarce in yellow catfish.

SSRs are tandem repeating sequences of 1–6 nucleotides and distributed throughout vertebrate genomes [8]. Based on their locations, SSRs can be classified into genomic SSRs (gSSRs) and Expressed Sequence Tag-SSRs (EST-SSRs) [9]. Because of high level of polymorphism, SSRs have wide applications in population genetics, such as parentage analysis [10], Quantitative Trait Locus (QTL) mapping [11], marker assisted selection (MAS) [12], and phylogenetic studies [13]. Traditional methods of developing gSSR markers require fragmented genomic DNA and are usually time-consuming and labor-intensive. With the advent of high-throughput sequencing technology, the development of EST-SSRs has become a fast, efficient, and low-cost option for economical fish species [14,15].

The transcriptome of yellow catfish was acquired using a 454 GS-FLX Titanium platform and 540 Mbp of raw data were generated. In this study, we analyze the frequency and distribution of 82,794 potential EST-SSRs in the yellow catfish transcriptome. Sixty of 300 validated primer pairs were selected and further characterized for polymorphism analysis. Recently, we have performed genetic selection breeding on four wild populations of yellow catfish collected from Chang Lake (Jingzhou), Hong Lake (Honghu), South Lake (Zhongxiang) and Dongting Lake (Hunan) as previously reported [16]. These EST-SSR markers should provide a promising genetic resource for molecular breeding of yellow catfish.

2. Results and Discussion

2.1. Characterization of EST-SSRs in the Yellow Catfish Transcriptome

Putative open reading frames (ORFs) of all the assembled contigs and singletons were predicted by EMBOSS software. After analyzing the transcriptome by MISA software, we identified 82,794 SSRs, among which 23,085 SSRs (27.9%) are located in the coding region, 18,954 SSRs (22.9%) in the 5'-UTR, and 18,537 SSRs (22.4%) in the 3'-UTR (Figure 1A). Then, we analyzed the distribution of
SSRs that have 2–6 bp repeat motif and are widely used. Of the 14,090 SSR identified in the coding region, dinucleotide accounts for 72.2% (10,180), tri-nucleotide is 17.6% (2478), tetra-nucleotide is 9.3% (1309), followed by penta-nucleotide 0.7% (98) and hexa-nucleotide 0.2% (25). Of the 10,584 SSR identified in the 5'-UTR, the most abundant is also dinucleotide accounting for 74.3% (7868), followed by tri-, tetra-, penta- and hexa-nucleotide with 14.5% (1532), 10% (1061), 1.1% (118) and 0.04% (5), respectively. Of the 11,654 SSR in the 3'-UTR, the percentage (and number) of di-, tri-, tetra-, penta- and hexa-nucleotide is 77.4% (9015), 13.4% (1559), 8.2% (961), 0.9% (107) and 0.1% (12), respectively (Figure 1B). Different locations of SSR markers in ESTs may suggest their possible for gene expression and functions [17]. The SSR insertions inside the promoter region of genes could modulate their expression levels [18].

**Figure 1.** Distribution of EST-SSRs across the 5' UTR, CDS and 3' UTR in yellow catfish. Number of SSRs located on non-coding and coding region (A) and the distributions of SSRs with different motif sizes (B).

Among the 82,794 SSRs, di-nucleotide is the most abundant type of repeat motif that is accounting for 65.14% (53,933) of the total SSRs, while hexa-nucleotide is the least type (84, 0.10%). Furthermore, the percentages of mono-, tri-, tetra-, and penta-nucleotide are 17.11% (14,168), 9.79% (8104), 7.28% (6027) and 0.58% (478) in respective. Most of SSRs had 6–36 repeat units, and six repeat units (15,004, 18.12%) and ten repeat units (9784, 11.82%) were the most represented types (Table 1). In the di-nucleotide repeat SSRs, AC/GT (39,554, 73.3%) and AG/CT (11,460, 21.2%) are the dominant types (Figure 2A). Similar to other fishes [19], (GC)n repeats are extremely rare in yellow catfish. Two most frequent repeats in the tri- nucleotide are AAT/ATT (3645, 45.0%) and ATC/GAT (1353, 16.7%) (Figure 2B). Among the tetra- nucleotide, the top two types of repeat motifs are AAAT/ATTT (1412, 23.4%) and ACAG/CTGT (943, 15.6%) (Figure 2C).
Table 1. Frequency of different repeat motifs among the EST-SSRs of yellow catfish.

| Repeats | Mo | Di | Tri | Tetra | Penta | Hexa | Total | Percentage (%) |
|---------|----|----|-----|-------|-------|------|-------|----------------|
| 5       | -  | 0  | 2654| 1843  | 253   | 43   | 4793  | 5.79           |
| 6       | -  | 12,561| 1347| 994   | 80    | 22   | 15,004| 18.12          |
| 7       | -  | 7110 | 893 | 632   | 44    | 8    | 8687  | 10.49          |
| 8       | -  | 4411 | 537 | 421   | 16    | 5    | 5390  | 6.51           |
| 9       | -  | 3248 | 384 | 316   | 18    | 3    | 3969  | 4.79           |
| 10      | 6769| 2429| 276 | 289   | 19    | 2    | 9784  | 11.82          |
| 11      | 3055| 1972| 263 | 225   | 15    | 0    | 5530  | 6.68           |
| 12      | 1805| 1628| 244 | 194   | 4     | 1    | 3876  | 4.68           |
| 13      | 995 | 1418| 207 | 144   | 14    | 0    | 2778  | 3.36           |
| 14      | 602 | 1260| 206 | 129   | 6     | 0    | 2203  | 2.66           |
| 15      | 392 | 1112| 173 | 132   | 2     | 0    | 1811  | 2.19           |
| 16      | 174 | 1008| 186 | 96    | 2     | 0    | 1466  | 1.77           |
| 17      | 136 | 896 | 141 | 110   | 1     | 0    | 1284  | 1.55           |
| 18      | 80  | 846 | 113 | 64    | 0     | 0    | 1103  | 1.33           |
| 19      | 53  | 806 | 128 | 60    | 3     | 0    | 1050  | 1.27           |
| 20      | 26  | 799 | 90  | 46    | 1     | 0    | 962   | 1.16           |
| 21      | 18  | 731 | 81  | 58    | 0     | 0    | 888   | 1.07           |
| 22      | 13  | 688 | 54  | 44    | 0     | 0    | 799   | 0.97           |
| 23      | 12  | 713 | 44  | 48    | 0     | 0    | 817   | 0.99           |
| 24      | 5   | 709 | 30  | 26    | 0     | 0    | 770   | 0.93           |
| 25      | 3   | 655 | 23  | 30    | 0     | 0    | 711   | 0.86           |
| 26      | 4   | 634 | 12  | 23    | 0     | 0    | 673   | 0.81           |
| 27      | 1   | 648 | 9   | 20    | 0     | 0    | 678   | 0.82           |
| 28      | 3   | 573 | 3   | 12    | 0     | 0    | 591   | 0.71           |
| 29      | 0   | 594 | 1   | 12    | 0     | 0    | 607   | 0.73           |
| 30      | 3   | 563 | 1   | 12    | 0     | 0    | 579   | 0.70           |
| 31      | 5   | 521 | 0   | 6     | 0     | 0    | 532   | 0.64           |
| 32      | 2   | 479 | 2   | 7     | 0     | 0    | 490   | 0.59           |
| 33      | 0   | 462 | 2   | 2     | 0     | 0    | 466   | 0.56           |
| 34      | 0   | 432 | 0   | 3     | 0     | 0    | 435   | 0.53           |
| 35      | 1   | 421 | 0   | 5     | 0     | 0    | 427   | 0.52           |
| 36      | 0   | 394 | 0   | 5     | 0     | 0    | 399   | 0.48           |
| >36     | 11  | 3212| 0   | 19    | 0     | 0    | 3242  | 3.92           |
| Total   | 14,168| 53,933| 8104| 6027 | 478   | 84   | 82,794| 100.00         |
| Percentage (%) | 17.11 | 65.14 | 9.79 | 7.28 | 0.58  | 0.10 | 100.00 |             |

2.2. SSR Marker Development and Genetic Diversity Analysis

A total of 300 SSR primers located on 280 assembled contigs and singletons were randomly selected and amplified using DNA templates extracted from four wild populations of yellow catfish from Chang Lake, Hong Lake, South Lake and Dongting Lake. Of these SSR primers, 263 (87.7%) pairs of primers exhibited stable and repeatable amplification, and 57 (19%) of them were identified as polymorphic loci in all 48 individuals. Although we tried multiple PCR reactions under different amplification conditions, the 37 pair of primers still did not produce any PCR fragment, which probably due to assembly errors in sequences or primer pairs designed across a splice site with a large intron [20]. Among the 263 worked and 37 not-worked SSRs, there are 122 (46.4%) and 11 (29.7%)
SSRs in the 3'-UTR, 71 (27.0%) and 12 (32.4%) SSRs in the 5'-UTR, 66 (25.1%) and 13 (35.1%) SSRs in the coding region, respectively. Further, there are 106 polymorphic and 157 unpolymorphic SSR markers, in which 41 (38.7%) and 81 (51.6%), 33 (31.1%) and 38 (24.2%), 30 (28.3%) and 36 (22.9%) SSRs were respectively located in the 3'-UTR, 5'-UTR and coding region. Moreover, tetra-nucleotide repeat is the most frequent form in both polymorphic SSRs (67.0%, 24 in the 3'-UTR, 21 in the 5'-UTR and 26 in the coding region) and unpolymorphic SSRs (51.6%, 36 in the 3'-UTR, 22 in the 5'-UTR and 23 in the coding region).

Figure 2. Characterization and frequency of different motifs among dinucleotide repeats (A), trinucleotide repeats (B) and the tetranucleotide repeats (C) EST-SSRs of yellow catfish.

A representative set of yellow catfish accessions amplified by primer pair H86 was shown in Figure 3. The selected 57 polymorphic primer pair sequences were characterized and deposited in GenBank to provide a foundation for breeding and genetic research of yellow catfish (Table 2).

Across the four populations of 48 individuals surveyed, the number of alleles ($N_a$) per locus varied widely among the markers (Table 2) and ranged from 2 to 17, with an average of 8.23 alleles. We made an analysis of the observed ($H_o$) and expected heterozygosity ($H_e$). The former value was ranged from 0.04 to 1.00 with an average of 0.52, while the latter varied from 0.12 to 0.92 with an average of 0.70. The high value of mean $H_o$ and $H_e$ suggests that there is a relatively high heterozygosity. The overall polymorphic index content (PIC) values were ranged from 0.12 to 0.91 with an average of 0.66. According to the criterion previously described, three categories were defined as high (PIC > 0.5), moderate (0.25 < PIC < 0.5) and low (PIC < 0.25) [21,22]. So these 57 primers exhibited high levels of PIC. Lastly, the fixation index ($F_{IS}$) values were ranged from $-0.83$ to 0.93 with an average of 0.25.
Table 2. Characteristics of the 57 EST-SSR markers for yellow catfish. Population genetic diversity analysis at 57 SSR loci was shown under the parameters: number of alleles per locus (\(N_a\)), observed heterozygosity (\(H_o\)), expected heterozygosity (\(H_e\)), polymorphic information content (\(PIC\)) and fixation index (\(FIS\)).

| EST-SSR | Repeat Motif | Primer Sequences (5′–3′) | \(T_a\) (°C) | Allele Size Range (bp) | Description of Putative Function | GenBank Accession No. | \(N_a\) | \(H_o\) | \(H_e\) | \(PIC\) | \(FIS\) |
|---------|--------------|---------------------------|--------------|------------------------|---------------------------------|----------------------|--------|--------|--------|-------|--------|
| H2      | (AAT)13      | F: CTTCCAGGGGCTTCATAAGT    | 51           | 138–180                | F-box and WD repeat containing   | KM211716             | 7      | 0.604  | 0.831  | 0.80  | 0.266  |
|         |              | R: GTGGTGTCGTCGTCGTCG     |              |                        | protein 7                       |                      |        |        |        |       |        |
| H6      | (ATAG)16     | F: TGTTGTAATCTCTCTATAGAAGG | 53           | 252–348                | Transposable element Tc1         | KM216910             | 13     | 0.729  | 0.865  | 0.84  | 0.148  |
|         |              | R: TGTTGGAGAAACATAGAAGTTGA|              |                        | transposase                      |                      |        |        |        |       |        |
| H13     | (GT)10       | F: AGAGCTAGGCCAAACTGCTG    | 53           | 141–205                | Calcium binding protein 39       | KM236563             | 7      | 0.917  | 0.720  | 0.67  | −0.286 |
|         |              | R: TCAGGAAGAACAAAGCTGG     |              |                        |                                 |                      |        |        |        |       |        |
| H15     | (CA)15       | F: CTCGACCAATCTCTGAGTCTTTC| 53           | 209–240                | NF-kappa-B inhibitor beta        | KM216912             | 5      | 0.271  | 0.565  | 0.47  | 0.515  |
|         |              | R: GTCATCATCAACGGACAACG    |              |                        |                                 |                      |        |        |        |       |        |
| H16     | (CA)17       | F: GAGAGACAGCGAGCCTCAGT    | 58           | 121–180                | NEDD4-like E3 ubiquitin protein  | KM216871             | 16     | 1.000  | 0.924  | 0.91  | −0.094 |
|         |              | R: CTAGGGCACACACCTCCT      |              |                        | ligase WWP2                      |                      |        |        |        |       |        |
| H17     | (TTA)14      | F: ACCACCTCAGCACACGC       | 57           | 110–172                | Hypothetical protein             | KM216905             | 7      | 0.500  | 0.815  | 0.78  | 0.380  |
|         |              | R: CACACCTTCTAAATGAACATCA  |              |                        |                                 |                      |        |        |        |       |        |
| H20     | (TTA)17      | F: ATGGTGTTCTCCACAGTGCGAG | 58           | 152–248                | No significant match             | KM216903             | 11     | 0.542  | 0.824  | 0.80  | 0.336  |
|         |              | R: CCGTCGTTGACCAGATGTT     |              |                        |                                 |                      |        |        |        |       |        |
| H28     | (TGGAGC)6    | F: GGGGCCTCTCCTGTTATTTA    | 57           | 153–216                | Gonadal-soma derived growth      | KM216886             | 7      | 0.375  | 0.725  | 0.68  | 0.477  |
|         |              | R: GTGCAGCGCTTGGAAACTAGG   |              |                        | factor precursor                 |                      |        |        |        |       |        |
| H29     | (TTTTA)17    | F: GCCCTACACGACACGTCAGAC   | 57           | 102–132                | Protein regulator of cytokinesis  | KM216864             | 4      | 0.417  | 0.550  | 0.47  | 0.234  |
|         |              | R: CGAGCAAGATCTCCTCACC     |              |                        | 1a                                |                      |        |        |        |       |        |
| H32     | (TGATGT)8    | F: TTGGGGGTTAAAAATGATCCCG  | 58           | 197–345                | Predicted protein                | KM216901             | 10     | 0.500  | 0.774  | 0.74  | 0.347  |
|         |              | R: CGAGAAGGCGTTAAAAAGGG    |              |                        |                                 |                      |        |        |        |       |        |
| H66     | (AG)7        | F: ATGGGATGACCAGAGGAGCAG   | 59           | 263–300                | cAMP-dependent protein kinase     | KM236564             | 3      | 0.083  | 0.120  | 0.12  | 0.299  |
|         |              | R: GTCTCTTCTGTTGCTCG       |              |                        | catalytic subunit beta           |                      |        |        |        |       |        |
| EST-SSR | Repeat Motif | Primer Sequences (5'–3') | Ta (°C) | Allele Size Range (bp) | Description of Putative Function | GenBank Accession No. | Heterozygosity |
|---------|--------------|--------------------------|--------|------------------------|----------------------------------|----------------------|---------------|
|         |              |                          |        |                        |                                  |                      | Na  | Ho  | He  | PIC | FIS |
| H77     | (TG)7        | F: AAGCATAGATTGTGCGGTCT R: TCAGCTGTAGGCCATGTC | 58     | 264–334                | Glucocorticoid receptor 2        | KM216888             | 3   | 0.354 | 0.298 | 0.26  | −0.201 |
| H78     | (GTAT)9      | F: GACCAAAGTGATCGGACTC R: AATAACCGACATCTCTGATC | 62     | 273–378                | Glucocorticoid receptor 2        | KM216909             | 3   | 1.00 | 0.552 | 0.44  | −0.829 |
| H84     | (AC)24       | F: TGTAAGGAAAAGGACACACACA R: GTGAGGGTGTTGCAAGGTT | 58     | 202–284                | Low density lipoprotein receptor | KM216916             | 7   | 1.00 | 0.837 | 0.81  | −0.207 |
| H86     | (TG)11tc(TG)8 | F: CTCTCCAGAGTGTTGTCGG R: GTGGTCGATACCAGAAGGA | 59     | 255–305                | Adenylate cyclase type 5         | KM216892             | 9   | 0.917 | 0.715 | 0.66  | −0.297 |
| H89     | (TGGA)5      | F: AATGACAATAGGGTGCGGAG R: TCATCCATCGTCCGATCCG | 59     | 269–339                | No significant match             | KM216896             | 3   | 0.208 | 0.194 | 0.18  | −0.085 |
| H96     | (GAAT)5      | F: GCACCTCGTCGAGGTATAG R: TACCAGCTGCTGATGTA | 59     | 173–181                | No significant match             | KM216857             | 2   | 0.292 | 0.252 | 0.22  | −0.171 |
| H106    | (TTCT)5      | F: TGATTTTTGGGACAGAGGAA R: TCAAACCTCAAGTCAGGAAGCAA | 59     | 202–264                | No significant match             | KM216856             | 14  | 0.604 | 0.903 | 0.88  | 0.324 |
| H107    | (TTCT)5      | F: TGATTTTTGGGACAGAGGAA R: TCAAACCTCAAGTCAGGAAGCAA | 58     | 238–294                | No significant match             | KM216891             | 5   | 0.375 | 0.622 | 0.56  | 0.391 |
| H109    | (TTTG)6      | F: TATTTTTCTGTGGCTCTCTG R: TACAGAACTGAGTGGTGTG | 58     | 275–315                | Heterogeneous nuclear ribonucleoprotein U protein 1 | KM216875 | 13 | 0.417 | 0.908 | 0.89  | 0.537 |
| H114    | (TCTG)5      | F: TGAGGGGGTGGTATTCTTTTG R: GGAGGAACGAAGAACGACAC | 59     | 215–322                | Probable palmitoyltransferase ZDHHC20-like | KM216914 | 5  | 0.313 | 0.636 | 0.57  | 0.503 |
| H135    | (ATCTA)5     | F: GCATGACATGTCTGTGGTT R: TGAAGGTCAGGCTGAGCAA | 59     | 140–225                | No significant match             | KM216858             | 9   | 0.563 | 0.737 | 0.69  | 0.229 |
| H139    | (TTAGC)6     | F: GCCGCCATGGCTAGCTGCC R: CAAAACCCACGACACTCG | 58     | 154–204                | Cyclin-dependent kinase 2 associated protein 2 | KM216895 | 4   | 0.042 | 0.609 | 0.52  | 0.931 |
| H147    | (TCTA)25     | F: TTGCCCAATTATACACCATTGTT R: TCAGCATTTAAATGGAGGCCAC | 58     | 229–264                | Uncharacterized protein LOC101056656, partial | KM216859 | 14  | 0.563 | 0.818 | 0.79  | 0.305 |
## Table 2. Cont.

| EST-SSR | Repeat Motif | Primer Sequences (5'–3') | T<sub>a</sub> (°C) | Allele Size Range (bp) | Description of Putative Function | GenBank Accession No. | Heterozygosity |
|---------|--------------|----------------------------|---------------------|------------------------|---------------------------------|----------------------|---------------|
|         |              | F: TTGCACCTATTGGGGGATGTTG | 58                  | 210–272                | Hypothetical protein            | KM216860             | 11            |
|         |              | R: AACGGGAGGCCTCTAAACCGAGT |                     |                        | PANDA_009670                   |                      | 0.604 | 0.790 | 0.76 | 0.227 |
| H151    | (TGTT)11     | F: CACTGATGATGGAATTGGGA   | 59                  | 143–183                | Glycogen phosphorylase,         | KM216904             | 5             |
|         |              | R: TCCCCGCTCTGACAGTTTT    |                     |                        | liver form                      |                      | 0.438 | 0.711 | 0.65 | 0.378 |
| H152    | (AGTT)15     | F: GAAACGGATATTAGTGGGGG   | 59                  | 191–252                | No significant match            | KM216879             | 10            |
|         |              | R: GCAAATCCCAATAGAGCGAA  |                     |                        |                                |                      | 0.771 | 0.868 | 0.84 | 0.102 |
| H153    | (ACAT)12     | F: TGCCAGTACTGACAAACCA    | 58                  | 164–204                | Collagen type IV alpha-3-binding | KM216898             | 8             |
|         |              | R: TTTTTAGTGGCCCATGTTT    |                     |                        | protein-like                    |                      | 0.625 | 0.762 | 0.72 | 0.172 |
| H154    | (TTTC)14     | F: GAACGTCTCTTGTCTTGC     | 58                  | 223–283                | E3 ubiquitin-protein ligase MIB2| KM216861             | 17            |
|         |              | R: GTAGGGACTGACGATGGGAA   |                     |                        |                                |                      | 0.604 | 0.924 | 0.91 | 0.339 |
| H155    | (AATA)15     | F: CTTTCATTTGTGCGTTGGC    | 59                  | 232–344                | No significant match            | KM216862             | 11            |
|         |              | R: GGACATCGTAGCCAACTTCC   |                     |                        |                                |                      | 0.604 | 0.857 | 0.83 | 0.288 |
| H156    | (AAAT)15     | F: CATAACCGCACTGAATATGTA | 58                  | 211–259                | Family with sequence similarity | KM216885             | 7             |
|         |              | R: AGCTGTATTTCTAAAGGCGGA |                     |                        | 222, member B                   |                      | 0.521 | 0.801 | 0.77 | 0.343 |
| H158    | (ATTT)16     | F: ATCCATGCATCCTCACACA    | 60                  | 223–307                | No significant match            | KM216894             | 6             |
|         |              | R: ACATCTGCATTGGGATATCC   |                     |                        |                                |                      | 0.500 | 0.753 | 0.71 | 0.329 |
| H159    | (ATCT)22     | F: TTTCACTATTGGTGTATCC    | 58                  | 217–332                | No significant match            | KM216893             | 4             |
|         |              | R: TCCTCAACGAGGTTAGTAC    |                     |                        |                                |                      | 0.271 | 0.613 | 0.55 | 0.554 |
| H160    | (TTCT)11     | F: CGTTGCACATTGTTGGTTTA  | 59                  | 217–278                | No significant match            | KM216865             | 14            |
|         |              | R: TGGAGTTGCAAACATGAGGCC |                     |                        |                                |                      | 0.417 | 0.751 | 0.73 | 0.440 |
| H161    | (CCAT)11     | F: AGCAACAGTCGGAGGACATA  | 59                  | 161–202                | Hypothetical protein            | KM216854             | 8             |
|         |              | R: TGGTTGGGTTGATAGATGTT  |                     |                        | PANDA_019388                   |                      | 0.792 | 0.779 | 0.74 | −0.027 |
| H163    | (AAAT)11     | F: GCTTTGATCACTTCTTCC    | 58                  | 286–382                | No significant match            | KM216884             | 4             |
|         |              | R: TGTTGCTAGCCCATGCTCGAA |                     |                        |                                |                      | 0.583 | 0.659 | 0.59 | 0.106 |
| H165    | (CACT)11     | F: GCGGAGACCGCTCTGTGATC  | 58                  | 171–255                | Muscle creatine kinase          | KM216887             | 9             |
|         |              | R: AGGATGCAGCTGATTCAAGTC |                     |                        |                                |                      | 0.583 | 0.823 | 0.79 | 0.284 |
| EST-SSR | Repeat Motif | Primer Sequences (5’–3’) | Tₐ (°C) | Allele Size Range (bp) | Description of Putative Function | GenBank Accession No. | Heterozygosity |
|---------|--------------|---------------------------|---------|------------------------|---------------------------------|----------------------|---------------|
|         |              |                           |         |                        |                                 |                      | Nₜ  | H₀  | Hₑ  | PIC | FIS |
| H166    | (TGTT)₁₁     | F: AGCGTTAGCTTAGCACTG    | 58      | 157–233                | Hypothetical protein            | KM216899             | 14  | 0.729 | 0.838 | 0.81 | 0.121 |
|         |              | R: ACACACACACCGAGCATG    |         |                        | ZEAMMB73_428483                |                      |     |       |       |     |     |
| H168    | (ATCC)₁₀     | F: TGATCAGTGACCTAGAGC    | 58      | 258–334                | No significant match            | KM216863             | 5   | 0.417 | 0.537 | 0.46 | 0.216 |
|         |              | R: TGATCAGTGACCTAGAGC    |         |                        |                                 |                      |     |       |       |     |     |
| H169    | (CATC)₁₁     | F: CGATCAGTAGCTACTCCTCC  | 58      | 221–292                | Rho GTPase-activating protein    | KM216906             | 7   | 0.563 | 0.805 | 0.77 | 0.294 |
|         |              | R: CATGACTTGCCACCCCTAGTA |         |                        | 7-like                          |                      |     |       |       |     |     |
| H171    | (ATAC)₁₀     | F: GATTCACCCAAAATGACATG  | 58      | 173–248                | Tribbles homolog 3              | KM216872             | 10  | 0.271 | 0.492 | 0.48 | 0.444 |
|         |              | R: AAAGGCAATGACACTGCTCC  |         |                        |                                 |                      |     |       |       |     |     |
| H172    | (AGAA)₁₀     | F: AGTGGTTCCGTTGGAGGTTT  | 58      | 255–328                | No significant match            | KM216913             | 6   | 0.500 | 0.762 | 0.72 | 0.337 |
|         |              | R: TTCTGACGTCTTCATGCTGC  |         |                        |                                 |                      |     |       |       |     |     |
| H176    | (AATA)₁₀     | F: TGAAGGTCAGAAATGACAGC  | 58      | 118–145                | No significant match            | KM216876             | 5   | 0.833 | 0.761 | 0.71 | −0.107 |
|         |              | R: CTGACCAGAAACAGCTGAA  |         |                        |                                 |                      |     |       |       |     |     |
| H203    | (TGAT)₈      | F: CAGGCGGCGTTTTCTTTTC   | 58      | 131–157                | Protein LBH-like                | KM216869             | 9   | 0.521 | 0.786 | 0.75 | 0.330 |
|         |              | R: CAGAACCGCTGTGCTTTTA  |         |                        |                                 |                      |     |       |       |     |     |
| H216    | (CTTT)₈      | F: GATGATGAGGTGTAGAGGC   | 58      | 113–151                | No significant match            | KM216874             | 6   | 0.625 | 0.729 | 0.69 | 0.134 |
|         |              | R: TTTTTGACGTACACAGAGTCA |         |                        |                                 |                      |     |       |       |     |     |
| H217    | (ATTT)₈      | F: CTGAAAAGGAAAAACACATCTG| 58      | 231–257                | No significant match            | KM216908             | 5   | 0.458 | 0.656 | 0.59 | 0.294 |
|         |              | R: TTTCCAGTGTCACACGTACGA|         |                        |                                 |                      |     |       |       |     |     |
| H228    | (TTTA)₈      | F: CCGAGACGCTTAAAGGACTTG | 61      | 204–272                | Zgc:63767 protein               | KM216915             | 12  | 0.354 | 0.835 | 0.81 | 0.572 |
|         |              | R: GCTACAGATCAGAGCAGCGTGTC |         |                        |                                 |                      |     |       |       |     |     |
| H229    | (ATTT)₈      | F: TTTTGCAAAACGAATATCACCA| 58      | 197–252                | No significant match            | KM216907             | 11  | 0.479 | 0.765 | 0.74 | 0.367 |
|         |              | R: CCCCCACAAAAACCTGTGTAAAT |         |                        |                                 |                      |     |       |       |     |     |
| H233    | (ATCA)₈      | F: CCACCTGAAAAGCTCAGAAC  | 58      | 244–286                | No significant match            | KM216890             | 8   | 0.229 | 0.497 | 0.47 | 0.534 |
|         |              | R: TACGTCGTTCCACACGAGAG |         |                        |                                 |                      |     |       |       |     |     |
| H237    | (CTTT)₈      | F: TGGGTAGTGCTGTTGGTCACG | 58      | 248–301                | No significant match            | KM216880             | 12  | 0.458 | 0.841 | 0.82 | 0.449 |
### Table 2. Cont.

| EST-SSR | Repeat Motif | Primer Sequences (5’–3’) | T (°C)  | Allele Size Range (bp) | Description of Putative Function | GenBank Accession No. | Heterozygosity |
|---------|--------------|--------------------------|---------|------------------------|----------------------------------|-----------------------|---------------|
|         |              |                          |         |                        |                                  |                       | Na | Ho | He | PIC | FIS  |
| H246 (ATA)9 | F: GACGCAGCTCGTGAATGTTR R: AACCCTCACAATACTCCCACAC | 58 | 223–294 | No significant match | KM216883 | 10 | 0.625 | 0.821 | 0.79 | 0.230 |
| H249 (ATT)13 | F: GGGGAAATAGTTATGAAAATGGR R: CACTCGCCTCTCTAAAAGCAC | 58 | 276–326 | No significant match | KM216877 | 9 | 0.229 | 0.684 | 0.62 | 0.662 |
| H251 (AATG)9 | F: CTGAGATAGGCACAGGCTCC R: ACCCCGTTCCAGTGTTCTC | 58 | 244–324 | C1orf43-like protein | KM216866 | 9 | 0.375 | 0.656 | 0.63 | 0.423 |
| H254 (ATAA)8 | F: TTCACTCAATTTCTGTTTCAA R: TGTGGGGTATGCGTACGAC | 58 | 282–319 | No significant match | KM216870 | 7 | 0.646 | 0.685 | 0.64 | 0.048 |
| H256 (GAAT)8 | F: CAATGCAACAGCATGGAGG R: CTGAGGTGCAAACGTGAT | 58 | 212–346 | No significant match | KM216902 | 15 | 0.792 | 0.879 | 0.86 | 0.090 |
| H259 (ATTT)12 | F: CAGCATGGCTTTTTTGTTR R: GGTTCATGAGCAACTCCAA | 56 | 263–326 | No significant match | KM216853 | 8 | 0.333 | 0.613 | 0.59 | 0.451 |
| H260 (TCTG)17 | F: GGATGGGAGGTTTGGAA R: TCAGTCTCCATTACACTGGG | 58 | 218–248 | No significant match | KM216855 | 6 | 0.208 | 0.620 | 0.55 | 0.660 |
Figure 3. PCR amplification profiles of 48 yellow catfish accessions using primer pair H86. The PCR amplified products were separated on 7% polyacrylamide gel. M indicated the molecular markers.

3. Experimental Section

3.1. Fish Samples

Four wild populations of yellow catfish (2–3 years old) were collected from Chang Lake (Jingzhou), Hong Lake (Honghu), South Lake (Zhongxiang) and Dongting Lake (Hunan), as described previously [16]. 12 individuals were randomly selected from each population. Experimental protocols used here were approved by the institution animal care and use committee of Huazhong Agricultural University.

3.2. SSR Identification and Development of Primer Pairs

We have carried out 454 pyrosequencing technology to perform high-throughput deep sequencing of the yellow catfish transcriptome, with a cDNA library constructed by one RNA pool which has an equal quantity of total RNA extracted from ovary, testis, liver, kidney, muscle, brain, spleen and heart of yellow catfish (accession number of NCBI archive database: SRP032172). All types of SSRs from dinucleotides to hexanucleotides were identified from the assembled contigs and singletons using MISA software under default parameter settings: a minimum of ten repeats for dinucleotide SSRs, six repeats for dinucleotide SSRs, five repeats for trinucleotide, tetranucleotide pentanucleotide and hexanucleotide SSRs. Then we designed primers for the microsatellite sequences using the software Primer Premier 5.0.

3.3. Genomic DNA Extraction, PCR Amplification and Electrophoresis

Genomic DNA was extracted from the tail fin following the traditional proteinase K and phenol-chloroform extraction method, as described by Wang et al. [1]. The concentration of DNA was adjusted to 100 ng/μL, and DNA was stored at −20 °C until used.
To initially evaluate the polymorphism of the identified microsatellite markers, polymerase chain reaction (PCR) was performed using a 10 μL total volume that contained 0.5 mM each primer, 0.25 μL each dNTP, 0.25 μL PCR buffer, 1 μL MgCl2, 0.5 units of Taq polymerase, and approximate 50 ng DNA. The following conditions were used for the PCR: 1 cycle of denaturation at 95 °C for 5 min and 35 cycles of 30 s at 94 °C, 30 s at a primer-specific annealing temperature, and 45 s at 72 °C. In the final step, the products were extended for 7 min at 72 °C. The PCR products were separated on 7% native polyacrylamide gel and visualized via silver staining. The allele size was estimated according to the pUC18 marker (TianGen Biotech, Beijing, China).

3.4. Evaluation of SSR Polymorphism and Genetic Diversity Analysis

To determine the polymorphism of these SSR loci, optimized primers were used to perform PCR reaction with genomic DNA extracted from 48 individuals of these four populations. PCR amplification was performed to accurately screen population-level variation, and PCR products were subjected to electrophoresis 7.0% non-denaturing polyacrylamide gels. To test the level of polymorphism at each EST–SSR locus in four populations, the number of observed alleles (N_a), observed heterozygosities (H_o) and expected heterozygosities (H_e), fixation index (F_IS) and polymorphism information content (P_IC) values were calculated using POPGENE (Version 1.31) and CERVUS (Version 3.0.3).

4. Conclusions

By exploiting 454 transcriptome sequencing database, we obtained much information of EST-SSR makers. We not only developed 57 available EST-SSR makers, but also evaluated the population genetics of wild yellow catfish. This is the first report of a comprehensive study on the development and analysis of SSR markers by high-throughput sequencing in yellow catfish. Our results will provide a set of available EST-SSR markers that will be essential for future molecular breeding and genetic studies of yellow catfish.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31301931), the Fundamental Research Funds for the Central Universities (52902-0900202496, 2013PY068), the National Key Basic Research Program (2010CB126301) and the special Fund for Agro-scientific Research in the Public Interest from the Ministry of Agriculture of China (2009030406). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: Jin Zhang, Jie Mei and Jian-Fang Gui. Performed the experiments: Jin Zhang, Wenge Ma, Xiaomin Song, Qiaohong Lin. Bioinformatics analysis and wrote the manuscript: Jin Zhang, Jie Mei, and Jian-Fang Gui. All authors read and approved the final paper.

Conflicts of Interest

The authors declare no conflict of interest.
References

1. Wang, Z.W.; Zhu, H.P.; Wang, D.; Jiang, F.F.; Guo, W.; Zhou, L.; Gui, J.F. A novel nucleo-cytoplasmic hybrid clone formed via androgenesis in polyploid gibel carp. *BMC Res. Notes* 2011, 4, doi:10.1186/1756-0500-4-82.

2. Gutierrez, A.P.; Lubieniecki, K.P.; Fukui, S.; Withler, R.E.; Swift, B.; Davidson, W.S. Detection of quantitative trait loci (QTL) related to grilseing and late sexual maturation in Atlantic salmon (*Salmo salar*). *Mar. Biotechnol.* 2014, 16, 103–110.

3. Wang, D.; Mao, H.L.; Chen, H.X.; Liu, H.Q.; Gui, J.F. Isolation of Y- and X-linked SCAR markers in yellow catfish and application in the production of all-male populations. *Anim. Genet.* 2009, 40, 978–981.

4. Kumla, S.; Doolgindachbaporn, S.; Sudmoon, R.; Sattayasai, N. Genetic variation, population structure and identification of yellow catfish, *Mystus nemurus* (C&V) in Thailand using RAPD, ISSR and SCAR marker. *Mol. Biol. Rep.* 2012, 39, 5201–5210.

5. Fishery Bureau of Ministry of Agriculture PRC. *China Fishery Statistical Yearbook*; China Agriculture Press: Beijing, China, 2010.

6. Gui, J.; Zhu, Z. Molecular basis and genetic improvement of economically important traits in aquaculture animals. *Chin. Sci. Bull.* 2012, 57, 1751–1760.

7. Mei, J.; Gui, J.F. Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. *Sci. Chin. Life Sci.* 2014, 57, in press.

8. Toth, G.; Gaspari, Z.; Jurka, J. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* 2000, 10, 967–981.

9. Chung, J.W.; Kim, T.S.; Suresh, S.; Lee, S.Y.; Cho, G.T. Development of 65 novel polymorphic cDNA-SSR markers in common vetch (*Vicia sativa* subsp. *sativa*) using next generation sequencing. *Molecules* 2013, 18, 8376–8392.

10. Poetsch, M.; Bahnisch, E.; Ludescher, F.; Dammann, P. Maximising the power of discrimination is important in microsatellite-based paternity analysis in songbirds. *J. Ornithol.* 2012, 153, 873–880.

11. Keong, B.P.; Siraj, S.S.; Daud, S.K.; Panandam, J.M.; Rahman, A.N.A. Identification of quantitative trait locus (QTL) linked to dorsal fin length from preliminary linkage map of molly fish, *Poecilia* sp. *Gene* 2014, 536, 114–117.

12. Song, W.T.; Li, Y.Z.; Zhao, Y.W.; Liu, Y.; Niu, Y.Z.; Pang, R.Y.; Miao, G.D.; Liao, X.L.; Shao, C.W.; Gao, F.T.; *et al.* Construction of a High-Density Microsatellite Genetic Linkage Map and Mapping of Sexual and Growth-Related Traits in Half-Smooth Tongue Sole (*Cynoglossus semilaevis*). *PLoS One* 2012, 7, doi:10.1371/journal.pone.0052097.

13. Jia, X.D.; Wang, T.; Zhai, M.; Li, Y.R.; Guo, Z.R. Genetic diversity and identification of Chinese-grown pecan using ISSR and SSR markers. *Molecules* 2011, 16, 10078–10092.

14. Ribas, L.; Pardo, B.G.; Fernandez, C.; Alvarez-Dios, J.A.; Gomez-Tato, A.; Quiroga, M.I.; Planas, J.V.; Sitja-Bobadilla, A.; Martinez, P.; Piferrer, F. A combined strategy involving Sanger and 454 pyrosequencing increases genomic resources to aid in the management of reproduction, disease control and genetic selection in the turbot (*Scophthalmus maximus*). *BMC Genomics* 2013, 14, doi:10.1186/1471-2164-14-180.
15. Wang, J.; Yu, X.; Zhao, K.; Zhang, Y.; Tong, J.; Peng, Z. Microsatellite Development for an Endangered Bream Megalobrama pellegrini (Teleostei, Cyprinidae) Using 454 Sequencing. *Int. J. Mol. Sci.* **2012**, *13*, 3009–3021.

16. Dan, C.; Mei, J.; Wang, D.; Gui, J.F. Genetic differentiation and efficient sex-specific marker development of a pair of Y- and X-linked markers in yellow catfish. *Int. J. Biol. Sci.* **2013**, *9*, 1043–1049.

17. Lawson, M.J.; Zhang, L. Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region. *Gene* **2008**, *407*, 54–62.

18. Fuganti, R.; Machado Mde, F.; Lopes, V.S.; Silva, J.F.; Arias, C.A.; Marin, S.R.; Binneck, E.; Abdelnoor, R.V.; Marcelino, F.C.; Nepomuceno, A.L. Size of AT(n) insertions in promoter region modulates Gmhspl17.6-L mRNA transcript levels. *J. Biomed. Biotechnol.* **2010**, doi:10.1155/2010/847673.

19. Nagpure, N.S.; Rashid, I.; Pati, R.; Pathak, A.K.; Singh, M.; Singh, S.P.; Sarkar, U.K. FishMicrosat: A microsatellite database of commercially important fishes and shellfishes of the Indian subcontinent. *BMC Genomics* **2013**, *14*, 630, doi:10.1186/1471-2164-14-630.

20. Dutta, S.; Kumawat, G.; Singh, B.P.; Gupta, D.K.; Singh, S.; Dogra, V.; Gaikwad, K.; Sharma, T.R.; Raje, R.S.; Bandhopadhyya, T.K.; et al. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus cajan (L.) Millspaugh]. *BMC Plant. Biol.* **2011**, *11*, 17, doi:10.1186/1471-2229-11-17.

21. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.

22. Yadav, H.K.; Ranjan, A.; Asif, M.H.; Mantri, S.; Sawant, S.V.; Tuli, R. EST-derived SSR markers in *Jatropha curcas* L.: Development, characterization, polymorphism, and transferability across the species/genera. *Tree Genet. Genomes* **2010**, *7*, 207–219.

*Sample Availability*: All samples are available from the authors.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).